A general Cholesky decomposition based modeling of longitudinal IRT data.

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

Abstract

In this work we proposed an approach for modeling longitudinal Item Response Theory (IRT) data based on the work of Pourahmadi (1999), which uses the Cholesky decomposition of the matrix of variance and covariance (dependence) of interest which, in our case, is related to the latent traits. One of the most important features of this approach is that it handle unbalanced data (inclusion and dropouts of subjects) easily. Also, our modeling can accommodate various covariance (dependence) structures relatively easily, facilitates the choice of prior distributions for the parameters of the dependence matrix, facilitates the implementation of estimation algorithms, allows to consider different distributions for latent traits, makes it easier the inclusion of regression (growth curves) and multilevel structures for the latent traits, among other advantages. We focus on dichotomous responses, symmetric normal latent trait distributions and a single group of individuals followed over several evaluation conditions (time-points). In each of these evaluation conditions the subjects are submitted to a measuring instrument, that can be different along the time-points but have some structure of common items. For inference purposes, an appropriate augmented data structure is considered. 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) algorithm is used. Simulation studies indicate that the parameters are well recovered. In addition, a longitudinal study in education, promoted by the Brazilian federal government, is analyzed to illustrate the proposed methodologies.

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

1 Introduction

Longitudinal data are characterized when experimental units are followed along different measurement occasions (or time-points) that can not be mutually randomized. Along these measurement occasions, characteristics of interest realted to those experimental units are measured. In the IRT context the main interest lies on the response of subjects to items belonging to some measurement instrument (cognitive tests, psychiatric questionnaires, educational tests, among others) along different occasions (as scholar grades). These measurement instruments, in each time point, can be different from each other, but they must present some structure of common items. Due to this nested structure, that

---

[*]Department of Statistics, University of Campinas, Brazil
[†]Department of Statistics, University of Campinas, Brazil

is, the time-specific measurements within-subjects, it is expected to observe some within-subject dependence. On the LIRT (longitudinal IRT) data, the within subject responses to the items are assumed to be conditionally independent given the item parameters and the latent traits, whereas it is expected that some dependence structure be observed for the within-subject latent traits, see Andrade and Tavares (2005), Azevedo et al. (2012).

Some longitudinal item response theory models have been proposed. For example, Conaway (1990) proposed a Rasch model to analyze panel data using the marginal maximum likelihood approach, see Bock and Aitkin (1981). Eid (1996) defined a longitudinal model for confirmatory factorial analysis for polytomous response data. Andrade and Tavares (2005) and Tavares and Andrade (2006) developed an IRT model to estimate ability distribution parameters, considering several covariance structures, when item parameters are know and unknown, respectively. Recently, Azevedo et al. (2016) proposed a general class of IRT longitudinal models with multivariate normal distributions for the latent traits, considering a Bayesian framework. This class take into account important features of the longitudinal data, as time-heterogeneous latent trait variances and serial correlation. Furthermore, the authors proposed some Bayesian model fit assessment tools. However, the multivariate distribution considered to model the latent traits, make the developed MCMC algorithms (and even the obtaining of maximum likelihood estimates) computationally cumbersome, when the number of subjects and the number of measurement occasions are large. In addition, even though the authors explored some structured covariance matrices, their approach to estimate them involves a non structured full matrix. That is, despite of the restricted correlation pattern of the dependence matrix, the number of parameters to be estimated are not reduced and they increase when the number of measurement occasions increase. In this sense, it can also be difficulty to handle unbalanced data, induced by dropouts and/or inclusion of subjects during the study, for example. Indeed, in this sense, the authors considered only the balanced case. In addition, they did not consider the guessing parameter in the item response function.

Our goal is to develop a general Cholesky decomposition based modeling of longitudinal IRT data. To accomplish for that, the Antedependence Models, see Pourahmadi (1999) and Nunez-Anton and Zimmerman (2000) were considered. This approach is very flexible and allows for handling multivariate distributions through univariate conditional distributions. This feature can reduce the computational cost in MCMC algorithms, compared to the multivariate approach. It also allows to represent properly a wide range of specific correlation patterns, without considering additional random effects as in Azevedo (2008). In addition, it is quite useful to develop diagnostic tools as that ones residuals-based and based on posterior predictive techniques. With respect to prior specifications, the antedependence modeling is quite interesting, since it allows to define more flexible priors for structured covariance matrices. Furthermore it can easily handle unbalanced data and different latent trait distributions. Also, regression structures for the latent traits, as growth curve models and for the item parameters, as differential item functioning, are more easily accommodate. All of these features can be considered as advantages of our approach, compared with the previous works.

Concerning to the item response function (IRF), most of the previous works did not the consider the effect of guessing on the probability of response, when modeling response for multiple choice items. This effect is important to improve the estimation of the latent traits, specially those related to subjects with low latent trait values. In this work we consider the three-parameter probit model (Baker and Kim, 2004). However, other IRF's can be also considered and properly accommodated by using the MCMC algorithms

developed here through suitable augmented data structures.

This work is outlined as follows. In Section 2, the IRT longitudinal antedependence model is presented along with some of its properties. In Section 3, we present the prior and posterior distributions as well as all steps of our MCMC algorithms. In Section 4 some simulation studies are presented to study the efficiency of our model and MCMC algorithm, in terms of parameter recovery. In Section 5 a real data set from a Brazilian school development program is analyzed and some model fit assessment tools are presented, which were used in this data analysis. Finally, in the Section 6 we presented some conclusions and remarks.

2   Modeling

The IRT data structure consists in T time-points which one with n_t subjects ($t = 1, 2, \ldots, T$), answering tests with I_t items. Common items are defined across the tests, and it can be recognized as an incomplete block design. Then, the total number of items is $I \leq \sum_{t=1}^{T} I_t$ and the total number of latent traits is $n = \sum_{t=1}^{T} n_t$ and the total number of subjects is n^*. Dropouts and inclusions of subjects during the study are allowed. Let us define the following notation: θ_{jt} is the latent trait of subject j ($j = 1, 2, \ldots, n_t$) at the time-point t, $\boldsymbol{\theta}_{j.}$ is the latent traits vector of subject j, and $\boldsymbol{\theta}_{..}$ is the vector of all latent traits. Let Y_{ijt} be the response of subject j to the item i ($i = 1, 2, \ldots, I$) at the time-point t, $\mathbf{Y}_{.jt} = (Y_{1jt}, \ldots, Y_{I_tjt})'$ is the response vector subject j at the time-point t, $\mathbf{Y}_{...} = (\mathbf{Y}'_{.11}, \ldots, \mathbf{Y}'_{.n_TT})'$ is the entire response matrix and $(y_{ijt}, \mathbf{y}'_{.jt}, \mathbf{y}'_{...})'$ are the respective observed values. Let $\boldsymbol{\zeta}_i$ be the vector of item parameters of item i, $\boldsymbol{\zeta}$ the vector of all item parameters and $\boldsymbol{\eta_\theta}$ the vector of population parameters, related to the latent trait distribution.

Our IRT longitudinal model is defined in two levels: the level of responses and the level of the latent traits. In the first level is considered a three-parameter probit IRT model, which is suitable for dichotomous or dichotomized responses. In the second level we assume some appropriate multivariate distribution, that is

$$Y_{ijt}|\theta_{jt}, \boldsymbol{\zeta}_i \quad \sim \quad \text{Bernoulli}(P_{ijt}), \tag{1}$$
$$P_{ijt} = \mathbb{P}(Y_{ijt}|\theta_{jt}, \boldsymbol{\zeta}_i) \quad = \quad c_i + (1 - c_i)\Phi\left(a_i\theta_{jt} - b_i\right). \tag{2}$$

$$\boldsymbol{\theta}_{j.}|\boldsymbol{\eta_\theta} \sim D_T(\boldsymbol{\eta_\theta}), \tag{3}$$

where $D(.)$ stands for some T-variate distribution indexed by the parameters $\boldsymbol{\eta_\theta}$. In Equation (2), a_i denotes 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 covariance structures. The identification of a stuitable 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 the LIRT 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 and it allows to represent a wide range of the variance-covariance structures.

2.1 Antedependence models

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 $\mathbb{E}(\boldsymbol{\theta}_{j.}) = \boldsymbol{\mu}_\theta$ and $\mathrm{Cov}(\boldsymbol{\theta}_{j.}) = \boldsymbol{\Sigma}_\theta$.

Then, we can write the latent trait of the subject j $(j = 1, \ldots, n_t)$ at the time-point t as:

$$\theta_{jt} = \mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}) + \varepsilon_{jt}, \quad t = 1, 2, \ldots, T. \tag{4}$$

where ϕ_{tk} are the so-called generalized autoregressive parameters, see Pourahmadi (1999).

In matrix form, we have:

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

This model was named by Zimmerman and Núñez-Antón (2009) unstructured antedependence model. The random vectors $\boldsymbol{\varepsilon}_{j.} = (\varepsilon_{j1}, \varepsilon_{j2}, \ldots, \varepsilon_{jT})'$ are uncorrelated with $\mathrm{Cov}(\boldsymbol{\varepsilon}_j) = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix $\mathrm{diag}(d_1, d_2, \ldots, d_T)$ and $\mathbf{L}$ is a $(T \times T)$ lower-triangular matrix having the following form

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

Then, from (5) and using the definition of $\mathbf{D}$ we have that,

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_{j.}) = \mathbf{L}\,\mathrm{Cov}(\boldsymbol{\theta}_{j.} - \boldsymbol{\mu}_\theta)\mathbf{L}' = \mathbf{L}\boldsymbol{\Sigma}_\theta\mathbf{L}' = \mathbf{D}. \tag{6}$$

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

More parsimonious models can be obtained by considering some specific correlation patterns. When the restricted covariance model is supported by the data, we can reduce, considerably, the number of parameters to be estimated and it can improve the model fit compared to the unstructured model. Furthermore, the unstructured pattern may not be appropriate in more complex situations as unbalanced data design, small sample sizes with respect to the number of subjects and items and many measurement occasions (or time-points), see Azevedo et al. (2016) and Jennrich and Schluchter (1986) for more details.

For example, consider $T = 3$ time-points and the following structured matrix:

$$\boldsymbol{\Sigma}_\theta = \begin{pmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1}\sigma_{\theta_2}\rho_\theta & \sigma_{\theta_1}\sigma_{\theta_3}\rho_\theta^2 \\ \sigma_{\theta_1}\sigma_{\theta_2}\rho_\theta & \sigma_{\theta_2}^2 & \sigma_{\theta_2}\sigma_{\theta_3}\rho_\theta \\ \sigma_{\theta_1}\sigma_{\theta_3}\rho_\theta^2 & \sigma_{\theta_2}\sigma_{\theta_3}\rho_\theta & \sigma_{\theta_3}^2 \end{pmatrix}$$

This is the *first-order autoregressive* matrix time-heteroscedasticity. The $\mathbf{L}$ and $\mathbf{D}$ matrices are given by:

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{\sigma_{\theta_2}}{\sigma_{\theta_1}}\rho_\theta & 1 & 0 \\ 0 & -\frac{\sigma_{\theta_3}}{\sigma_{\theta_2}}\rho_\theta & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{D} = \begin{pmatrix} \sigma_{\theta_1}^2 & 0 & 0 \\ 0 & \sigma_{\theta_2}^2(1-\rho_\theta^2) & 0 \\ 0 & 0 & \sigma_{\theta_3}^2(1-\rho_\theta^2) \end{pmatrix}$$

By induction and through Equation (5) we can obtain the following linear model:

$$\begin{aligned} \theta_{j1} - \mu_{\theta_1} &= \varepsilon_{j1}, \\ \theta_{jt} - \mu_{\theta_t} &= \frac{\sigma_{\theta_t}}{\sigma_{\theta_{t-1}}}\rho_\theta(\theta_{j(t-1)} - \mu_{\theta_{t-1}}) + \varepsilon_{jt}, \quad t = 2, \ldots, T. \end{aligned} \tag{7}$$

Note that the parameters $(d_1, d_2, \ldots, d_T, \phi_{21}, \phi_{31}, \phi_{32}, \ldots, \phi_{T(T-1)})'$ are an one-to-one mapping of the parameters o f interest $(\sigma_{\theta_1}^2, \sigma_{\theta_2}^2, \ldots, \sigma_{\theta_T}^2, \rho_\theta)'$. This results is convenient to specify flexible prior distributions for the covariance parameters and also to implement MCMC algorithms.

Mostly, the matrices $\mathbf{L}$ and $\mathbf{D}$ do not have recognizable form. Then, for more complex structured matrix it is very difficult or not possible to obtain a general expression for the antedependence model (as in the expression (7)). In such cases the Cholesky decomposition is obtained numerically. Table 1 presents some examples of structured matrices considered in this work.

In order to represent the multivariate structure of the latent traits (see Equation 3) we will assume in Equation (4) that $\varepsilon_{jt} \overset{i.i.d}{\sim} N(0, d_t)$, $t = 1, 2, \ldots, T$, or equivalently,

$$\boldsymbol{\varepsilon}_{j.} = \mathbf{L}(\boldsymbol{\theta}_{j.} - \boldsymbol{\mu}_\theta), \quad \boldsymbol{\varepsilon}_{j.} \overset{ind.}{\sim} N_T(\mathbf{0}, \mathbf{D}), \tag{8}$$

where $\mathbf{D}$ is the diagonal matrix described before. By the normal distribution properties and the property (6) we can see that:

$$\boldsymbol{\theta}_{j.} \sim N_T(\boldsymbol{\mu}_\theta, \mathbf{L}^{-1}\mathbf{D}(\mathbf{L}^{-1})') \equiv N_T(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta). \tag{9}$$

We are also interested in the marginal distribution of the latent traits for each time-point. In this case we have,

$$\theta_t \sim N(\mu_{\theta_t}, \sigma_{\theta_t}^2), \quad t = 1, 2, \ldots, T \tag{10}$$

where,

$$\begin{aligned} \sigma_{\theta_t}^2 &= d_1 \\ \sigma_{\theta_t}^2 &= d_t + \sum_{k=1}^{t-1} l_{tk}^2 d_t, \end{aligned} \tag{11}$$

where l_{tk} are entries of the $\mathbf{L}^{-1}$ matrix.

Table 1: Structured covariance matrices used in this work.

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

3 Bayesian Estimation and MCMC Algorithms

In order to make easier the implementation of the MCMC algorithms we used the augmented data approach proposed by Béguin and Glas (2001). This methodology consists on defining a vector of binary variables W_{ijt} such that

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

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

$$\begin{aligned} \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 1, \theta_{jt}, \boldsymbol{\zeta}_i) & \propto & \Phi(a_i\theta_{jt} - b_i) \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 1, \theta_{jt}, \boldsymbol{\zeta}_i) & \propto & c_i(1 - \Phi(a_i\theta_{jt} - b_i)) \\ \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 0, \theta_{jt}, \boldsymbol{\zeta}_i) & = & 0 \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 0, \theta_{jt}, \boldsymbol{\zeta}_i) & = & 1. \end{aligned} \tag{12}$$

Therefore the augmented variables $\mathbf{Z} = (Z_{111}, \ldots, Z_{1n_11}, \ldots, Z_{In_TT})'$, are given by

$$Z_{ijt} | (\theta_{jt}, \zeta_i, w_{ijt}) = \begin{cases} N(a_i\theta_{jt} - b_i, 1)\mathbb{I}_{(z_{ijt} \geq 0)}, & \text{if } w_{ijt} = 1, \\ N(a_i\theta_{jt} - b_i, 1)\mathbb{I}_{(z_{ijt} < 0)}, & \text{if } w_{ijt} = 0. \end{cases} \tag{13}$$

The original response can be represented by

$$Y_{ijt} = \mathbb{I}(Z_{ijt} > 0)\mathbb{I}(W_{ijt} = 1) + \mathbb{I}(Z_{ijt} \leq 0)\mathbb{I}(W_{ijt} = 0), \tag{14}$$

where, $\mathbb{I}$ denotes the indicator function. To handle incomplete block design, see Montgomery (2004), another indicator variable $\mathbf{I}$ is defined as:

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

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

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

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

$$\begin{aligned} L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \mathbf{z}_{...}, \mathbf{w}_{...}, \mathbf{y}_{...}) \quad \propto \quad & \prod_{t=1}^{T}\prod_{j=1}^{n_t}\prod_{i \in I_{jt}} \exp\left\{-.5(z_{ijt} - a_i\theta_{jt} + b_i)^2\right\} \mathbb{I}_{(z_{ijt}, w_{ijt})} \\ \times \quad & p(w_{ijt}|y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i), \end{aligned} \tag{15}$$

where $\mathbb{I}_{(Z_{ijt}, w_{ijt})}$ stands for the indicator function $\mathbb{I}_{(z_{ijt} < 0, w_{ijt} = 0)} + \mathbb{1}_{(z_{ijt} \geq 0, w_{ijt} = 1)}$ and I_{jt} is the set of items answered by the subject j at time-point t and $p(w_{ijt}|y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i)$ follows from (12).

3.1 Prior and posterior distributions

The joint prior distribution is assumed to be as

$$p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\eta) = \left\{ p(\theta_{j1}|\boldsymbol{\eta}_{\theta_1}) \prod_{t=2}^{T}\prod_{j=1}^{n_t} p(\theta_{jt}|\theta_{j(1:t-1)}, \boldsymbol{\eta}_{\theta_t}) \right\} \left\{ \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i|\boldsymbol{\eta}_\zeta) \right\} \left\{ \prod_{t=1}^{T} p(\boldsymbol{\eta}_{\theta_t}|\boldsymbol{\eta}_\eta) \right\}, \tag{16}$$

where the subscript $(1 : t - 1)$ denotes the preceding latent traits. The prior distributions of the latent traits are conditionally defined according to the model (4) and the assumption that $\varepsilon_{jt} \overset{i.i.d}{\sim} N(0, d_t), \quad t = 1, 2, \ldots, T$. For the item parameters we have:

$$p(\boldsymbol{\zeta}_{i(-c_i)}) \propto \exp\left[-.5(\boldsymbol{\zeta}_{i(-c_i)} - \boldsymbol{\mu}_\zeta)' \boldsymbol{\Psi}_\zeta^{-1}(\boldsymbol{\zeta}_{i(-c_i)} - \boldsymbol{\mu}_\zeta)\right] \mathbb{I}_{(a_i > 0)}. \tag{17}$$

and,

$$c_i \sim \text{Beta}(a_c, b_c). \tag{18}$$

where $\boldsymbol{\zeta}_{i(-c_i)} = (a_i, b_i)$ is independent of c_i. For the population parameters we have:

$$
\begin{aligned}
\mu_{\theta_t} &\sim N(m_\mu, \sigma_\mu^2), \\
d_t &\sim IG(a_d, b_d), \quad t = 1, \ldots, T, \\
\phi_{tk} &\sim N(\mu_\phi, \sigma_\phi^2) \quad t = 2, \ldots, T \text{ and } k = 1, \ldots, t-1.
\end{aligned} \tag{19}
$$

where IG denotes the inverse-gamma distribution with $\mathbb{E}(d_t) = a_d/b_d$ and $Var(d_t) = a_d/b_d^2$. For the structured covariance matrices (see Table 1) the prior distributions for correlation parameters are directly specified as:

$$
\rho_{\theta_t} \overset{i.i.d}{\sim} N(\mu_\rho, \sigma_\rho^2)\mathbb{I}_{[0,1]}, \quad t = 1, 2, \ldots, T-1. \tag{20}
$$

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

Given the augmented likelihood in Equation (15) and the prior distribution in Equations (4), (17), (18) and (19), the joint posterior distribution is given by:

$$
\begin{aligned}
p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, |\mathbf{z}_{...}, \mathbf{w}_{...}, \mathbf{y}_{...}) \propto & \left\{ \prod_{t=1}^{T} \prod_{j=1}^{n} \prod_{i \in I_{jt}} \exp\left\{-.5(z_{ijt} - a_i\theta_{jt} + b_i)^2\right\} p(w_{ijt}|y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i)\mathbb{I}_{(z_{ijt}, w_{ijt})} \right\} \\
& \times \left\{ \prod_{j=1}^{n_t} p(\theta_{j1}, \boldsymbol{\eta}_{\theta_t}) \prod_{t=2}^{T} p(\theta_{jt}|\theta_{j(1:t-1)}, \boldsymbol{\eta}_{\theta_t}) \right\} \left\{ \prod_{i=1}^{I} p(\boldsymbol{\zeta}_i|\boldsymbol{\eta}_\zeta) \right\} \\
& \times \left\{ \prod_{t=1}^{T} p(\boldsymbol{\eta}_{\theta_t}|\boldsymbol{\eta}_\eta) \right\},
\end{aligned} \tag{21}
$$

where $\boldsymbol{\eta}_\zeta$ and $\boldsymbol{\eta}_\eta$ are hyperparameters associated with $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}_\theta$, respectively. In addition, we are assuming independence between items and population parameters. This model is not identified, since differente values of the parameters can lead to the same likelihood. As pointed out by Tavares and Andrade (2006) and Azevedo et al. (2016), the identification can be achieved by fixing the mean and the variance of the latent traits distributions for any arbitrary time-point, for example $\mu_{\theta_1} = 0$ and $\psi_{\theta_1} = 1$ along with a suitable test design (as mentioned before and ahead). In our approach, these restrictions can be implemented easier compared to the appraaches of Azevedo et al. (2016) and Tavares and Andrade (2006). This is another advantage of our model. Since the posterior distribution has an intractable form, we will use MCMC algorithms in order to obtain empirical approximation for the posterior marginal distributions. We will show that the full conditional distribution of the model parameters $\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta$ are easy to sample from (for $\boldsymbol{\eta}_\theta$ this holds only for the unstructured matrix). For the latent traits we have:

$$
\theta_{jt}|(.) \sim N(b_t B_t, B_t) \text{ where,}
$$

$$
b_t = \begin{cases}
\sum_{i \in I_{jt}} a_i(z_{ijt} + b_i) + \mu_{\theta_t}\left(\frac{1}{d_t} + \frac{\phi_{(t+1)t}}{d_{t+1}}\right) + \frac{\phi_{(t+1)t}}{d_{t+1}}(\theta_{j(t+1)} - \mu_{\theta_{t+1}}), & t = 1 \\[2em]
\begin{aligned} &\sum_{i \in I_{jt}} a_i(z_{ijt} + b_i) + \frac{\phi_{(t+1)t}}{d_{t+1}}\left[(\theta_{j(t+1)} - \mu_{\theta_{t+1}}) - \sum_{k=1}^{t-1} \phi_{(t+1)k}(\theta_{jk} - \mu_{\theta_k})\right] \\ &\quad + \frac{1}{d_t}\left(\mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k})\right) + \frac{\mu_{\theta_t}\phi_{(t+1)t}^2}{d_{t+1}}, \end{aligned} & t = 2, \ldots, T-1 \\[2em]
\sum_{i \in I_{jt}} a_i(z_{ijt} + b_i) + \frac{1}{d_t}\left(\mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k})\right), & t = T
\end{cases}
$$

8

and

$$B_t = \begin{cases} \left(\sum_{i \in I_{jt}} a_i^2 + \frac{1}{d_t} + \frac{\phi_{(t+1)t}^2}{d_{t+1}} \right)^{-1}, & t = 1, \ldots, T-1 \\ \left(\sum_{i \in I_{jt}} a_i^2 + \frac{1}{d_t} \right)^{-1}, & t = T. \end{cases} \qquad (22)$$

For the items parameters we have:

$$\boldsymbol{\zeta}_{i(-c_i)}|(.) \sim N(\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}\widehat{\boldsymbol{\zeta}}_i, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}), \qquad (23)$$

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

(.) denotes the set of all necessary 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 guessing parameters,

$$c_i|(.) \sim \text{Beta}(s_i + a_c - 1, t_i - s_i + b_c - 1) \qquad (24)$$

where $s_i = \sum_{j|w_{ijt}=0}^{n} \mathbf{y}_{ij.}$ and $\sum_{j=1}^{n} \mathbb{I}(w_{ijt} = 0)$. For the population parameters we have the following full conditional distributions:

$$\mu_{\theta_t}|(.) \sim N(A_t a_t, A_t) \text{ where,} \qquad (25)$$

$$a_t = \frac{1}{d_t} \sum_{j=1}^{n_t} \left[\theta_{jt} - \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}) \right] + \frac{m_\mu}{\sigma_\mu^2} \text{ and } A_t = \left(\frac{n_t}{d_t} + \frac{1}{\sigma_\mu^2} \right)^{-1}.$$

$$d_t|(.) \sim IG(\frac{n_t}{2} + a_d, \frac{s_t}{2} + b_d) \text{ where,} \qquad (26)$$

$$s_t = \sum_{j=1}^{n_t} \left[\theta_{jt} - \mu_{\theta_t} - \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}) \right]^2, \quad t = 1, \ldots, T$$

.

$$\phi_{tk}|(.) \sim N(Q_{tk}q_{tk}, Q_{tk}) \text{ where,}$$

$$\begin{aligned} Q_{tk} &= \left(\frac{\sum_{j=1}^{n_t}(\theta_{j(t-1)} - \mu_{\theta_{t-1}})^2}{d_t} + \frac{1}{\sigma_\phi^2} \right)^{-1} \\ & \qquad (27) \\ q_{tk} &= \frac{1}{d_t} \sum_{j=1}^{n_t}(\theta_{jk} - \mu_{\theta_k})(\theta_{jt} - \mu_{\theta_t} - \sum_{k \neq t} \phi_{tk}(\theta_{jk} - \mu_{\theta_k})). \end{aligned}$$

for all $t = 2, \ldots, T$ and $k = 1, \ldots, (t-1)$. The marginal variances can be recovered by using the relation (11). Therefore, a full Gibbs sampling (FGS) algorithm to estimate the model parameters is definied in Algorithm 3.1.

The algorithm 3.1 is suitable to fit unstructured covariance matrices. For the structured matrices (see Table 1), we need to use an auxiliary algorithm to sampling from the correlation

Algorithm 3.1 Full Gibbs sampling algorithm

1: Start the algorithm by choosing suitable initial values. Repeat steps 2-9.
2: Simulate W_{ijt} from $W_{ijt}|(.)$ for all $i = 1, \ldots, I$, $j = 1, \ldots, n_t$ e $t = 1, \ldots, T$.
3: Simulate Z_{ijt} from $Z_{ijt}|(.)$ for all $i = 1, \ldots, I$, $j = 1, \ldots, n_t$ e $t = 1, \ldots, T$.
4: Simulate θ_{jt} from $\theta_{jt}|(.)$ for all $j = 1, \ldots, n_t$ e $t = 1, \ldots, T$.
5: Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i|(.)$ for all $i = 1, \ldots, I$.
6: Simulate c_i from $c_i|(.)$ for all $i = 1, \ldots, I$.
7: Simulate μ_{θ_t} from $\mu_{\theta_t}|(.)$ for all $t = 1, \ldots, T$.
8: Simulate d_t from $d_t|(.)$ for all $t = 1, \ldots, T$.
9: Simulate ϕ_{tk} from $\phi_{tk}|(.)$ for all $t = 2, \ldots, T$ and $k = 1, \ldots, t-1$.

---

parameters taking into account the restrictions of the covariance matrix. On the other hand, the latent traits simulation can also be improved if we consider a sampler that take into account the dependence structure imposed by the model. In next sections we will discuss two auxiliary algorithms which can handle these two issues.

3.2 A FFBS based sampler for the latent traits

In this section we focus on the latent trait estimation. As mentioned before, the latent traits are assumed to have an antedependence structure with errors normally distributed (see Equation 4). A common way to sample from the joint full conditional distribution of $\boldsymbol{\theta}_{j\cdot}$ is to consider univariate full conditional distributions, in order to sample from a time-specific latent trait given all others, through Gibbs sampling. It means that: $\theta_{jt}|\boldsymbol{\theta}_{j(-t)}$ for all $t = 1, \ldots, T$, where $\boldsymbol{\theta}_{j(-t)}$ is the latent traits vector without the t-th component. That is the idea of the full conditional distribution presented in Equation (22). However, this procedure can generate chains with high autocorrelations, specially in the presence of many time-points (Gamerman and Lopes, 2006). Carter and Kohn (1994) and Frühwirth-Schnatter (1994) have proposed a sampling scheme for *dynamic models*, which allows to sample the so-called state parameters jointly, based on the Kalman filter. It is called **F**orward **F**iltering **B**ackward **S**ampling. In the following we presented this approach adapted to our problem.

3.2.1 Dynamic Models

Dynamic models are defined by the pair of equations called *observation* and *system or evolution* equations. Using the notation of the Gamerman and Lopes (2006) we have:

$$
\begin{aligned}
u_t &= F_t^\top \theta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\
\theta_t &= G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, \Omega_t),
\end{aligned}
\tag{28}
$$

where u_t is a sequence of observations at the time, conditionally independent given θ_t and σ_t^2. The specification of the model is completed with the prior $\theta_1 \sim N(r, R)$. For our IRT model, considering the augmented data structure, we have the following representation in terms of dynamic models:

$$
\begin{aligned}
Z_{jit} &= a_i \theta_{jt} - b_i + \xi_{jit}, \quad \xi_{jit} \sim N(0, 1), \\
\theta_{jt} &= \mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}) + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim N(0, d_t).
\end{aligned}
$$

for all $i \in I_{jt}$, $j = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$.

3.2.2 FFBS Algorithm

The FFBS algorithm basically consists on two steps: *forward* and *backward*. The forward step is performed by the Kalman filter procedure described below. Following Gamerman and Lopes (2006) consider the conditional distribution $\theta_{j(t-1)}|\mathbf{z}_{j.}^{t-1} \sim N(m_{j(t-1)}, C_{j(t-1)})$, where $\mathbf{z}_{j.}^{t-1}$ refers to the information until $t-1$. The system equation can be written as $\theta_{jt}|\theta_{j(t-1)} \sim N(\mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}), d_t)$. By proprieties of the normal distribution, these specifications can be combined leading to the following conditional distribution:

$$\theta_{jt}|\mathbf{z}_{j.}^{t-1} \sim N(a_{jt}, R_{jt}) \text{ where,} \tag{30}$$

$$a_{jt} = \mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{tk}(m_{jk} - \mu_{\theta_k}) \quad \text{and} \quad R_{jt} = d_t + \sum_{k=1}^{t-1} \phi_{tk}^2 C_{jk}.$$

Thus,

$$\theta_{jt}|\mathbf{z}_{j.}^{t} \sim N(m_{jt}, C_{jt}) \text{ where,} \tag{31}$$

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

Equation (31) is referred in the literature as *Kalman Filter*. Therefore, the backward distributions are given by:

$$\theta_{jt}|\theta_{j(t+1)}, \mathbf{z}_{j.}^{t} \sim N(m_{\theta_t}, C_{\theta_t}) \text{ where,} \tag{32}$$

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

where $\alpha_{t+1} = \mu_{\theta_{t+1}} - \phi_{t+1,t}\mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{t+1,t}(\theta_{jk} - \mu_{\theta_k})$, for all $j = 1, \ldots, n_t$ and $t = 1, \ldots, T$. Then, a scheme to sample from the full conditional distribution of $\boldsymbol{\theta}_{j.}$ is presented in algorithm 3.2.

Algorithm 3.2 FFBS algorithm

1: Sample θ_{jT} from $\theta_{jT}|\mathbf{z}_{j.}^{T}$ and set $t = T - 1$.
2: Sample θ_{jt} from $\theta_{jt}|\theta_{j(t+1)}, \mathbf{z}_{j.}^{t}$.
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 $t = T$.

3.3 Oversampling in the Single-Variable Exchange algorithm

In Section 3.1 we presented a full Gibbs sampling algorithm that is useful for unstructured covariance matrix. When a specific structure is imposed, the full conditional distributions of the correlation parameters are difficult to obtain, unlike to the generalized autoregressive parameters ϕ_{tk}. In this section we proposed an auxiliary sampler for the correlation parameters for structured matrices.

An usual way for sampling from an intractable posterior distribution is to use the *Metropolis-Hastings* (M-H) algorithm, see Gamerman and Lopes (2006). It requires a proposal distribution (a common choice is a random walk centred on the previous result) to generate proposed values. This value is accepted or rejected according to a transition probability. Then, the acceptation rate of the proposed values will depend on the variance of the proposal distribution, such that, when it is too large most values are rejected, whereas when it is too small only small steps are taken and the chain does not mix properly, see Marsman (2014). An alternative to overcome this problem was prosed by Murray et al. (2012), see algorithm (3.3). The so called *Single-Variable Exchange* algorithm (SVE) differs, from the M-H algorithm, in the way to propose *candidates* values.

Specifically, consider the following posterior distribution:

$$p(\vartheta|\mathbf{x}) \propto p(\mathbf{x}|\vartheta)p(\vartheta), \tag{33}$$

where ϑ and $\mathbf{x}$ represent a parameter and a set of observations, respectively. Therefore, $p(\mathbf{x}|\vartheta)$ and $p(\vartheta)$ denote the likelihood and prior distributions, respectively. Then an scheme to sample from posterior distribution (33) is given by algorithm 3.3.

Algorithm 3.3 The Single-Variable Exchange algorithm

1: Draw $\vartheta^{(m)} \sim p(\vartheta)$
2: Draw $\mathbf{x}^{(m)} \sim p(\mathbf{x}|\vartheta^{(m)})$
3: Draw $u \sim U(0,1)$
4: **if** $(u < \pi(\vartheta^{(m-1)} \to \vartheta^{(m)}))$ **then**
5: $\quad \vartheta^{(m-1)} = \vartheta^{(m)}$
6: **end if**

---

where,

$$\pi(\vartheta^{(m-1)} \to \vartheta^{(m)}) = \min\left\{1, \frac{p(\mathbf{x}|\vartheta^{(m)})p(\mathbf{x}^{(m)}|\vartheta^{(m-1)})}{p(\mathbf{x}|\vartheta^{(m-1)})p(\mathbf{x}^{(m)}|\vartheta^{(m)})}\right\}.$$

and $\vartheta^{(m)}$, $\vartheta^{(m-1)}$ denote a proposed and current values, respectively, for all simulation $m = 1, \ldots, M$.

Although the prior distribution closely resembles the target one, the SVE algorithm tends to frequently generate transition kernels for which the acceptance probability is low. To improve the efficiency of the SVE algorithm, in the sense to concentrate more probability mass on transition kernels with high acceptance probability, Marsman (2014) has been proposed an oversampling approach. It consists on simulating a number of i.i.d proposal values, instead simulating a single one. Let $t(\mathbf{x})$ be a sufficient statistic for the parameter ϑ and $t(\mathbf{x}^{(m)})$ the corresponding sufficient statistic of the proposed value $\vartheta^{(m)}$. Let s be the number of proposed values each one with its own value for the sufficient statistic, we have to choose that whose statistic $t(\mathbf{x}^{(m)})$ is closest to the observed $t(\mathbf{x})$, see algorithm (3.4).

To illustrate the use of the SVE with oversampling to estimate correlation parameters in an antedependence framework, consider the ARH(1) matrix. Let $\boldsymbol{\theta}_{..}$ be the set of all latent traits,

---

Algorithm 3.4 The Single-Variable Exchange algorithm with Oversampling

---

1: **for** $s = 1$ to S **do**
2:    Draw $\vartheta_s^{(m)} \sim p(\vartheta)$
3:    Draw $\mathbf{x}_s^{(m)} \sim p(\mathbf{x}|\vartheta_s^{(m)})$
4:    Compute $t(\mathbf{x}_s^{(m)})$
5: **end for**
6: Choose the $\vartheta_s^{(m)}$ whose $t(\mathbf{x}_s^{(m)})$ is closest to $t(\mathbf{x})$
7: Draw $u \sim U(0, 1)$
8: **if** $(u < \pi(\vartheta^{(m-1)} \to \vartheta^{(m)}))$ **then**
9:    $\vartheta^{(m-1)} = \vartheta^{(m)}$
10: **end if**

---

as defined in Section 2 and $p(\boldsymbol{\theta}_{..}|\boldsymbol{\mu}_\theta, \boldsymbol{\phi}, \mathbf{d})$ be the likelihood generated by the antedependence model defined in Equation (4), that is

$$
\begin{aligned}
p(\boldsymbol{\theta}_{..}|\boldsymbol{\mu}_\theta, \boldsymbol{\phi}, \mathbf{d}) &= \prod_{j=1}^{n_t} p(\theta_{j1}|\mu_{\theta_1}, d_1) \prod_{t=2}^{T} p(\theta_{jt}|\mu_{\theta_t}, \boldsymbol{\phi}_t, d_t) \\
&\propto \prod_{j=1}^{n_t} \frac{1}{\sqrt{d_1}} \exp\left\{ -\frac{1}{2d_1}\left(\theta_{j1} - \mu_{\theta_1}\right)^2 \right\} \\
&\times \prod_{t=1}^{T} \frac{1}{\sqrt{d_t}} \exp\left\{ -\frac{1}{2d_t}\left[(\theta_{jt} - \mu_{\theta_t}) - \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}) \right]^2 \right\},
\end{aligned} \tag{34}
$$

where $\boldsymbol{\mu}_\theta = (\mu_{\theta_1}, \ldots, \mu_{\theta_T})'$, $\boldsymbol{\phi} = (\phi_{21}, \phi_{31}, \phi_{32}, \ldots, \phi_{T(T-1)})'$, $\mathbf{d} = (d_1, \ldots, d_T)'$ and $\boldsymbol{\phi}_t$ denotes the elements correspondents to the time-point t. The algorithm 3.5 presents the correlation parameters sampler for the ARH(1) model.

When the structured matrix present two or more correlation parameters, the algorithm 3.5 can be applied independently for each one of those parameters, by choosing suitable sufficient statistics. It is also possible to sample blocks of correlation parameters. This can be done by modifying line 10 of the algorithm 3.5 to allow the acceptance/rejection of the proposed values, jointly.

In summary, we can combine the full Gibbs sampling with the FFBS algorithm by replacing step 4 of the algorithm 3.1 by algorithm 3.2, in order to improve the latent traits simulation. In the same way, we can combine FGS with SVE by replacing step 9 of the algorithm 3.1 by a SVE procedure, in order to simulate correlation parameters of a structured matrix. The algorithms 3.6 and 3.7 summarized these options.

4   Simulation Studies

To study the performance of the FFBS procedure, we compared algorithms 3.1 and 3.6. That is, the Gibbs sampling with and without FFBS. For brevity, these algorithms will be referred as FGS and FFBS, respectively.

In order to make a fair comparison, the effective sample size (ESS) criterion was applied, see Sahu (2002). The ESS is defined for each parameter as the number of MCMC samples drawn, says M, divided by the parameter's autocorrelation time, $\gamma = 1 + 2\sum_{k=1}^{\infty} \rho_k$, where ρ_k is the autocorrelation at lag k. Values of ESS close to M indicate low autocorrelation. There are many alternatives to estimate γ using sample autocorrelations, see Roberts (1996). We will use a simple upper bound $\frac{1+\rho^*}{1-\rho^*}$ where $\rho^* = |\rho_1|$. In this example we consider a real data set concerning to a Brazilian school development study (see Section 5, for additional details). The

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

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

then

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

Algorithm 3.6 Gibbs sampling with FFBS sampler for unstructured matrix

---
1: Start the algorithm by choosing suitable initial values. Repeat steps 2-9.
2: Simulate W_{ijt} from $W_{ijt}|(.)$ for all $i = 1, \ldots, I$, $j = 1, \ldots, n_t$ e $t = 1, \ldots, T$.
3: Simulate Z_{ijt} from $Z_{ijt}|(.)$ for all $i = 1, \ldots, I$, $j = 1, \ldots, n_t$ e $t = 1, \ldots, T$.
4: Simulate θ_{jt} using algorithm 3.2 for all $j = 1, \ldots, n_t$ e $t = 1, \ldots, T$.
5: Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i|(.)$ for all $i = 1, \ldots, I$.
6: Simulate c_i from $c_i|(.)$ for all $i = 1, \ldots, I$.
7: Simulate μ_{θ_t} from $\mu_{\theta_t}|(.)$ for all $t = 1, \ldots, T$.
8: Simulate d_t from $d_t|(.)$ for all $t = 1, \ldots, T$.
9: Simulate ϕ_{tk} from $\phi_{tk}|(.)$ for all $t = 2, \ldots, T$ and $k = 1, \ldots, t-1$.

Algorithm 3.7 Gibbs sampling with FFBS and SVE samplers for structured matrices

---
1: Start the algorithm by choosing suitable initial values. Repeat steps 2-9.
2: Simulate W_{ijt} from $W_{ijt}|(.)$ for all $i = 1, \ldots, I$, $j = 1, \ldots, n_t$ e $t = 1, \ldots, T$.
3: Simulate Z_{ijt} from $Z_{ijt}|(.)$ for all $i = 1, \ldots, I$, $j = 1, \ldots, n_t$ e $t = 1, \ldots, T$.
4: Simulate θ_{jt} using algorithm 3.2 for all $j = 1, \ldots, n_t$ e $t = 1, \ldots, T$.
5: Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i|(.)$ for all $i = 1, \ldots, I$.
6: Simulate c_i from $c_i|(.)$ for all $i = 1, \ldots, I$.
7: Simulate μ_{θ_t} from $\mu_{\theta_t}|(.)$ for all $t = 1, \ldots, T$.
8: Simulate d_t from $d_t|(.)$ for all $t = 1, \ldots, T$.
9: Simulate ρ_{θ_t} using a SVE procedure for all necessary correlation parameters.

---

model 2 was fitted according to each algorithm considering an unstructured covariance matrix. Chains of 30,000 iterations were considered after discarding 1,000 initial iterations. The ESS was computed for each parameter, as well as, the ESS per hour (ESS/h). Table 2 presents the average value for each algorithm. We can see that the FFBS outperforms the FGS algorithm, considering both ESS and ESS/h.

Table 2: Effective sample size and effective sample size per hour for FGS and FFBS algorithms

	ESS	ESS/h
FFBS	7045.69	286.90
FGS	6753.43	275.56

Henceforth, we will only consider algorithms 3.6 and 3.7 for unstructured and structured matrices, respectively.

4.1 Parameter Recovery

4.1.1 Balanced data

In this section we study the performance of our model and the proposed estimation algorithm concerning the parameter recovery. We consider different structured covariance matrices according to Table 1. Responses of $n_t = 1500, t = 1, 2, ..., 6$ subjects, along the time-points were simulated according to the model defined by Equations (1) and (2), for each covariance matrix of the Table 1. The item parameters were fixed in the following intervals: $a_i \in [.7, 2.62]$, $b_i^* \in [-1.95, 4]$ and the guessing parameter c_i assume the values $(.20, .21, .22, .23, .24, .25)$. The values of the difficulty parameters were fixed in order to consider low, middle and high difficulty in the items, compared with the population means along the time-points. The discrimination parameters values were fixed in order to cover many situations (items with high difficulty and high discrimination, high difficulty and low discrimination, etc). The tests structure is described as follows:

- Test 1: 20 items;

- Test 2: Test 1 + 20 other items;

- Test 3: the last 20 items of test 2 + 20 other items;

- Test 4: the last 20 items of test 3 + 20 other items;

- Test 5: the last 20 items of test 4 + 20 other items;

- Test 6: the last 20 items of test 5 + 20 other items;

Therefore, we have a total of $I = 120$ items. The latent traits were simulated from model (4) considering: $\boldsymbol{\mu}_\theta = (.0, 1.0, 1.4, 2.0, 2.3, 2.5)'$ and $\boldsymbol{\sigma}_\theta^2 = (1.00, .77, .30, .38, .59, .80)'$. The time-point 1 was considered as the reference one. Then we fixed increasing values for the population means on the $(0, 1)$ scale (respective the mean and the variance of the reference time-point), implying a growing in the latent traits over the time. This is an expected behavior in educational longitudinal studies, for example, see Santos et al. (2013) and Azevedo et al. (2012). The population variances induce a decreasing and then an increasing in the time-heterogeneity. We fixed from moderate to high values for population correlations in order to have a similar pattern to that observed in the real data. The correlation values can be seen in Table 3.

Table 4 presents the hyperparameters for the adopted prior distributions. Let us remember that the mean and the variance of reference time-point are zero and one, respectively. For

Table 3: Simulation study: correlation parameters for each dependence structure

Matrix	ρ_{θ_1}	ρ_{θ_2}	ρ_{θ_3}	ρ_{θ_4}	ρ_{θ_5}
ARH(1)	.8				
ARMAH(1,1)	.8	.5			
HT	.9	.85	.8	.6	.5
AD	.9	.85	.8	.6	.5

the population averages we are assuming that they vary reasonably around the mean of the reference group, and similarly, for the population variances related to the reference group. For the correlation parameters we assume that all are positive with a slight concentration toward zero but allowing positive values with a reasonable probability. The discrimination parameters are assumed to vary reasonably around a satisfactory discrimination power (grater than .6) and for the difficulty parameter we assume a value around the mean of the reference time-point with a variance which allows values close to the population means of the other time-points with a reasonable probability.

Table 4: Hyperparameters for the prior distribution

$\boldsymbol{\mu}_\zeta$	$\boldsymbol{\Psi}_\zeta$	$(m_\mu; \sigma_\mu^2)$	(a_d, b_d)	$(\mu_\phi, \sigma_\phi^2)$	$(\mu_\rho, \sigma_\rho^2)$
(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, Geweke's statistics and autocorrelations were monitored. We generate three chains based on three different sets of starting values. The inspection of the results indicate that a burn of 1,000 iterations and collecting samples at spacing of 30th iterations, generating a total of 31,000 values was enough to have valid MCMC samples of size 1,000 for each parameter.

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

Tables from 5 to 7 present the results for the population parameters. The mean and variance of the first time-point were fixed in 0 and 1, respectively, in order to define the scale. This restriction along with the linking test design ensure that the model is identified and, therefore, that all estimates lie in the same scale.

We can see that, the posterior means are very close to the respective true values, as well as, the most of the 95% equi-tailed credibility intervals cover the respective true values. Only two credibility intervals of the correlation parameters of the AD model did not cover the respective true values. Since we are dealing with only one replica, this could be due to sampling fluctuations. Table 8 presents the results for the latent traits and item parameters. They indicate that our algorithm recovered all parameters properly. Note that for the guessing parameters, the correlation is small. This is expected since the true values have low variability. However, the other results indicate that estimates were close to the true values. Figure 1 presents 95% equi-tailed credibility intervals for the item parameters of the ARH(1) model. Since for the other covariance models the results were similar, we decide did not present them, for the sake of

simplicity. We can see that, the credibility intervals cover all the true values. In a general way, we conclude that the parameters were properly recovered by the model and the estimation process.

Table 5: Simulation study for balanced data: results for population means

		Parameters					
		μ_{θ_1}	μ_{θ_2}	μ_{θ_3}	μ_{θ_4}	μ_{θ_5}	μ_{θ_6}
ARH(1)	True value	.000	1.000	1.400	2.000	2.300	2.500
	Mean	.000	1.004	1.379	1.991	2.317	2.552
	SD	–	.090	.144	.195	.231	.250
	CI(95%)	–	[.786, 1.085]	[1.015, 1.472]	[1.444, 2.116]	[1.644, 2.466]	[1.792, 2.719]
ARMAH(1,1)	Mean	.000	.980	1.400	1.947	2.225	2.406
	SD	–	.040	.057	.079	.093	.101
	CI(95%)	–	[.885, 1.052]	[1.253, 1.491]	[1.731, 2.052]	[1.965, 2.349]	[2.140, 2.540]
AD	Mean	.000	.923	1.257	1.886	2.176	2.365
	SD	–	.039	.052	.080	.099	.108
	CI(95%)	–	[.847, .991]	[1.139, 1.346]	[1.651, 1.983]	[1.867, 2.289]	[2.026, 2.490]
HT	Mean	.000	1.011	1.440	2.142	2.457	2.681
	SD	–	.037	.041	.066	.086	.107
	CI(95%)	–	[.942, 1.084]	[1.364, 1.519]	[2.028, 2.271]	[2.311, 2.628]	[2.496, 2.894]

Table 6: Simulation study for balanced data: results for population variance

		Parameters					
		$\sigma_{\theta_1}^2$	$\sigma_{\theta_2}^2$	$\sigma_{\theta_3}^2$	$\sigma_{\theta_4}^2$	$\sigma_{\theta_5}^2$	$\sigma_{\theta_6}^2$
ARH(1)	True value	1.000	.770	.300	.380	.590	.800
	Mean	1.00	.795	.327	.496	.758	1.139
	SD	–	.071	.043	.068	.097	.148
	CI(95%)	–	[.608, .919]	[.216, .390]	[.296, .607]	[.500, .922]	[.732, 1.382]
ARMAH(1,1)	Mean	1.00	.703	.271	.363	.601	.854
	SD	–	.057	.027	.037	.058	.084
	CI(95%)	–	[.564, .809]	[.209, .320]	[.282, .427]	[.479, .707]	[.692, 1.020]
AD	Mean	1.000	.655	.291	.435	.644	.568
	SD	–	.055	.029	.049	.072	.062
	CI(95%)	–	[.514, .749]	[.213, .341]	[.294, .516]	[.454, .770]	[.407, .681]
HT	Mean	1.000	.883	.365	.505	1.001	1.120
	SD	–	.063	.049	.089	.234	.273
	CI(95%)	–	[.767, 1.014]	[.277, .462]	[.357, .678]	[.647, 1.494]	[.699, 1.678]

Table 7: Simulation study for balanced data: results for population correlation

		Parameters				
		ρ_{θ_1}	ρ_{θ_2}	ρ_{θ_3}	ρ_{θ_4}	ρ_{θ_5}
ARH(1)	True value	.800	-	-	-	-
	Mean	.803	-	-	-	-
	SD	.015	-	-	-	-
	CI(95%)	[.788, .820]	-	-	-	-
ARMAH(1,1)	True value	.800	.500	-	-	-
	Mean	.738	.493	-	-	-
	SD	.020	.013	-	-	-
	CI(95%)	[.695, .774]	[.466, .518]	-	-	-
AD	True value	.900	.850	.800	.600	.500
	Mean	.844	.837	.787	.711	.642
	SD	.006	.002	.005	.010	.027
	CI(95%)	[.822, .846]	[.830, .844]	[.761, .795]	[.659, .715]	[.551, .652]
HT	Mean	.885	.801	.726	.617	.532
	SD	.003	.007	.011	.015	.015
	CI(95%)	[.878, .890]	[.786, .812]	[.703, .742]	[.584, .637]	[.492, .545]

Table 8: Simulation study for balanced data: results for the estimated latent traits and item parameters.

	Parameter	Statistic				
		Corr	MBias	MABias	MVAR	MRMSE
ARH(1)	Latent trait	.967	-.010	.232	.130	.471
	Discrimination	.917	.064	.131	.023	.232
	Difficulty	.996	-.041	.091	.055	.263
	Guessing	.287	-.000	.025	.003	.063
ARMAH(1,1)	Latent trait	.959	.040	.257	.112	.474
	Discrimination	.922	-.005	.115	.022	.219
	Difficulty	.997	.048	.087	.021	.179
	Guessing	.356	.004	.026	.003	.063
AD	Latent trait	.972	.098	.231	.085	.415
	Discrimination	.946	.002	.104	.021	.198
	Difficulty	.996	.110	.132	.022	.216
	Guessing	.292	.002	.024	.003	.061
HT	Latent trait	.972	-.088	.245	.098	.443
	Discrimination	.911	.157	.179	.021	.273
	Difficulty	.994	-.110	.169	.031	.283
	Guessing	.236	.003	.025	.003	.062

4.1.2 Unbalanced data

We simulated an unbalanced LIRT data, that is a longitudinal study with dropouts. More specifically, we simulated data considering the same scenario described in Subsection 4.1.1 and, for the sake of simplicity, we consider only the ARH(1) covariance matrix. The dropouts are related to the number of subjects per time, which were $n_1 = 1500, n_2 = 1400, n_3 = 1400, n_4 = 1400, n_5 = 1350$ and $n_6 = 1350$. Tables 9 and 10 present the results under unbalanced data. We can notice an increasing in the MABias and MRMSE compared with the results obtained for the balanced case. Also, population variances display higher posterior standard deviations compared with the balanced data case. Figures 1 and 2 present the estimates of the item parameters for the balanced and unbalanced case, respectively, under the ARH(1) structure. We can see that the estimates are similar, except for the discrimination parameters, which present a higher bias in the unbalanced case. We can see that population variances and discrimination parameters were more affected (in terms of the accuracy of the estimates) by the unbalanced design.

Table 9: Simulation study for unbalanced data: results for the population parameters

	True value	Mean	SD	CI(95%)
μ_{θ_1}	.000	.000	–	–
μ_{θ_2}	1.000	.896	.049	[.790, .991]
μ_{θ_3}	1.400	1.243	.060	[1.116, 1.343]
μ_{θ_4}	2.000	1.884	.088	[1.698, 2.021]
μ_{θ_5}	2.300	2.279	.116	[2.040, 2.467]
μ_{θ_6}	2.500	2.542	.131	[2.271, 2.750]
$\sigma^2_{\theta_1}$	1.000	1.000	–	–
$\sigma^2_{\theta_2}$	.770	1.491	.196	[1.185, 1.960]
$\sigma^2_{\theta_3}$	.300	.472	.053	[.375, .576]
$\sigma^2_{\theta_4}$	.380	.655	.075	[.497, .792]
$\sigma^2_{\theta_5}$	.590	1.356	.169	[1.012, 1.671]
$\sigma^2_{\theta_6}$	.800	1.580	.197	[1.146, 1.952]
ρ_θ	.800	.805	.008	[.790, .820]

Table 10: Simulation study for unbalanced data: results for the latent traits and item parameters

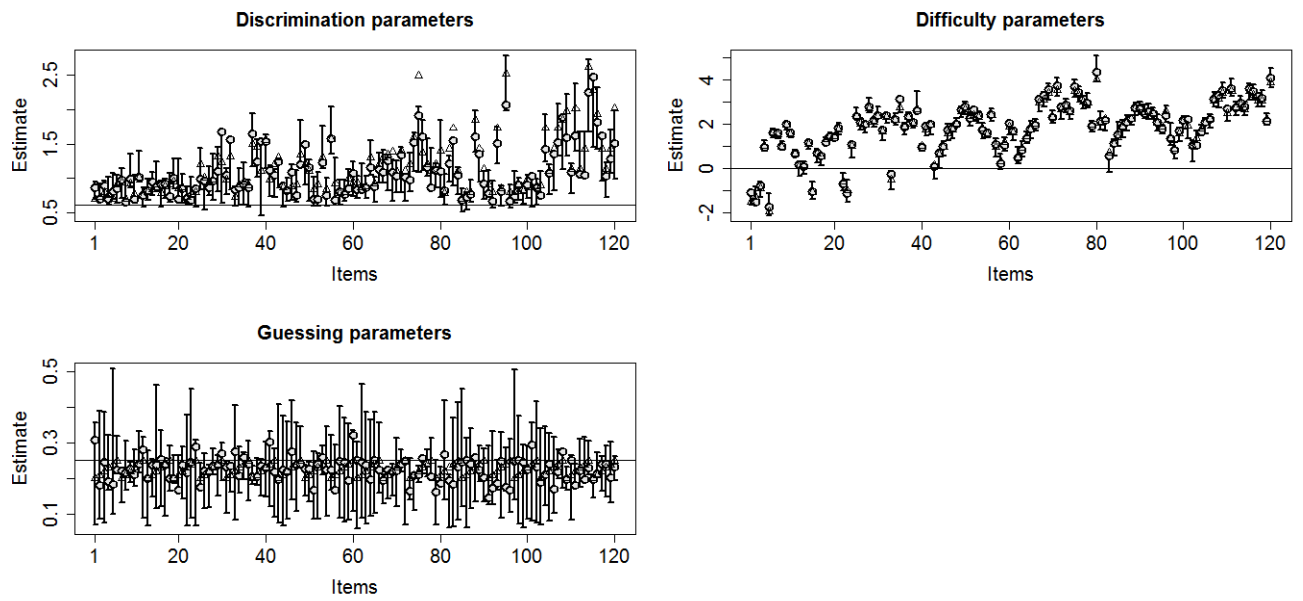|	Parameter	Statistic				
		Corr	MBias	MABias	MVAR	MRMSE
	Latent trait	.955	.058	.297	.135	.531
ARH(1)	Discrimination	.915	.248	.251	.015	.330
	Difficulty	.989	.047	.208	.037	.334
	Guessing	.353	.006	.026	.003	.064

Figure 1: Simulation study for balanced data: posterior means and 95% equi-tailed credibility intervals for item parameters under ARH(1) structure. Legend: Circles (estimates), triangles (true values) and Bars (credibility intervals)
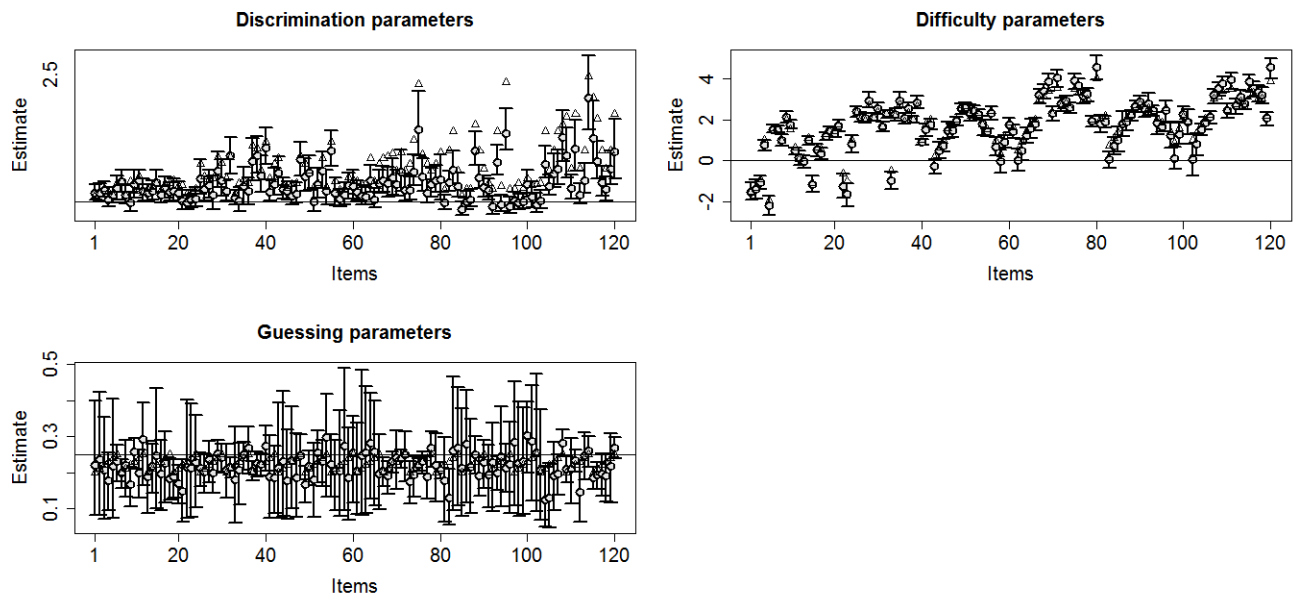

Figure 2: Simulation study for unbalanced data: posterior means and 95% equi-tailed credibility intervals for item parameters under ARH(1) structure. Legend: Circles (estimates), triangles (true values) and Bars (credibility intervals)

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 ideia it is to compare the observed and simulated data, where the latter is generated by using a suitable 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

$$p(\mathbf{y}_t^{rep}|\mathbf{y}_t^{obs}) = \int p(\mathbf{y}_t^{rep}|\boldsymbol{\vartheta}_t)p(\boldsymbol{\vartheta}_t|\mathbf{y}_t^{obs})d_{\boldsymbol{\vartheta}_t}, \tag{35}$$

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

$$\mathbb{P}(D(\mathbf{y}_t^{rep}|\boldsymbol{\vartheta}_t) \geq D(\mathbf{y}_t^{obs}|\boldsymbol{\vartheta}_t)|\mathbf{y}_t^{obs}) = \int_{D(\mathbf{y}_t^{rep}) \geq D(\mathbf{y}_t^{obs})} p(\mathbf{y}_t^{rep}|\mathbf{y}_t^{obs})d_{\mathbf{y}_t^{rep}}, \tag{36}$$

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 checking 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:

$$D_i = \sum_l \frac{|P_{li}^O - P_{li}^E|}{P_{li}^E}, \tag{37}$$

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 $l = 1, 2, \ldots, L$ and $i = 1, 2, \ldots, I$, 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 Spiegelhaltere 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 ρ_D statistics defined as $\overline{D(\vartheta)} - D(\bar{\vartheta})$. In our case we have,

$$D(\vartheta) = -2\text{Log}(L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}_{.}, \boldsymbol{\eta}_\theta)p(\theta|\boldsymbol{\eta}_\theta)). \tag{38}$$

where functions $L(.)$ and $p(.)$ are defined by equations (15) and (34). In practice, having M MCMC draws from the posterior distributions, the quantity $\overline{D(\vartheta)}$ can be estimated as:

$$\overline{D(\vartheta)} = \frac{1}{M} \sum_{m=1}^{M} D(\vartheta^{(m)}), \tag{39}$$

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

$$\widehat{\text{DIC}} = D(\bar{\vartheta}) + 2\rho_D, \tag{40}$$

$$\widehat{\text{EAIC}} = \overline{D(\vartheta)} + 2\rho_D, \tag{41}$$

$$\widehat{\text{EBIC}} - \overline{D(\vartheta)} + 2\log(n \times I), \tag{42}$$

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

5.3 Brazilian school development study

The analyzed data concern to a major study promoted by the Brazilian Federal Government know as the School Development Program. It aims to monitor the teaching quality in Brazilian public schools. A more detailed description of this data can be found in Azevedo et al. (2016). In a general way, it is a longitudinal study, performed to evaluate children's ability in Math and Portuguese language. Only the results concerning to Math part were considered in our analysis. A total of 1987 public school's students, selected from different regions of the country, were followed from fourth to eighth grade of the primary school, answering a different test in each one of these six different occasions, which are: 1999/April, 1999/November, 2000/November, 2001/November, 2002/November and 2003/November. A total of 167 items were considered in this analysis. Table 11 presents the structure of the tests, that is, the number of items per test and the number of common items across them.

Table 11: Structure of tests: real data analysis

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Test 1	34	10	5	1	0	0
Test 2	10	38	10	4	0	0
Test 3	5	10	36	7	3	1
Test 4	1	4	7	34	10	2
Test 5	0	0	3	10	40	10
Test 6	0	0	1	2	10	34

The antedependence IRT longitudinal model considering the unstructured covariance matrix, was applied to the data. The estimated covariance matrix can be seen in Equation (43). It presents the estimated variances on the main diagonal, estimated correlations above the main diagonal and estimated covariances below the main diagonal. The estimates indicate a time-heteroscedastic structure. Moreover, the correlations are high and decay slowly. Figure 3 compares the correlations estimated via unstructured model and the correlations estimated using the structured models considered in the previous section (see Table 1). We notice that the correlations estimated by ARH(1) and ARMAH(1,1) decay quickly and they are relatively different from those induced by the unstructured model, specially, for the fifth and sixth time-points. On the other hand the AD and HT models described suitably the correlation pattern, with a clear advantage of the AD model.

$$\begin{pmatrix} 1.000 & .848 & .738 & .695 & .677 & .590 \\ .835 & .970 & .816 & .749 & .732 & .649 \\ .626 & .683 & .721 & .818 & .770 & .693 \\ .646 & .686 & .646 & .866 & .846 & .741 \\ .679 & .724 & .657 & .790 & 1.008 & .824 \\ 1.087 & 1.177 & 1.084 & 1.272 & 1.525 & 3.397 \end{pmatrix} \tag{43}$$
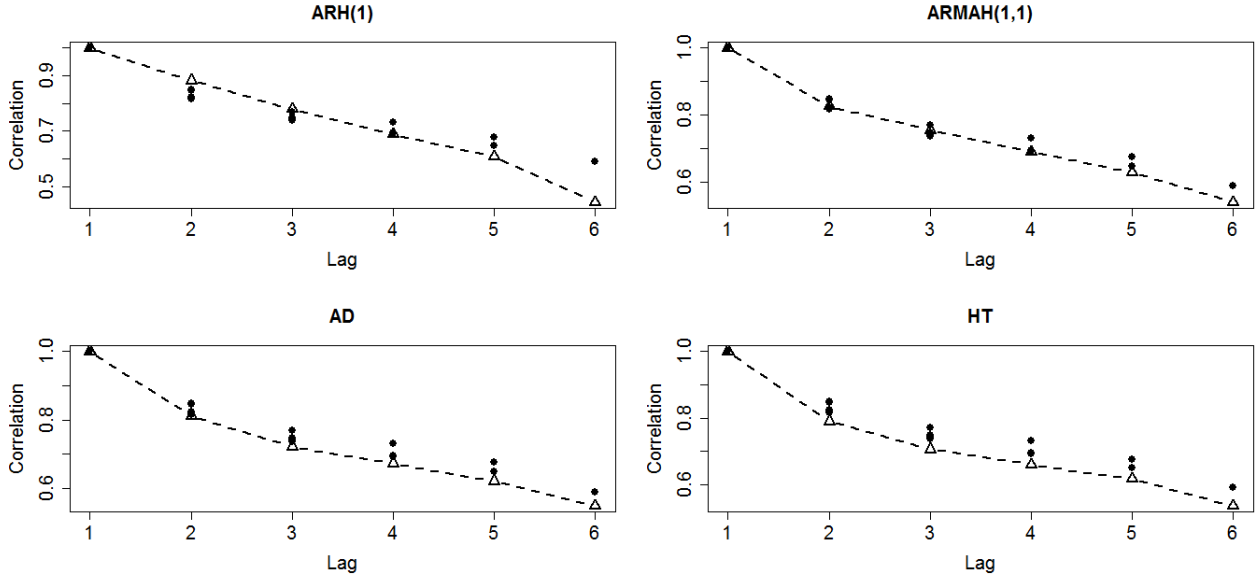
Figure 3: Correlation profiles. Legend: unstructured matrix (●), structured matrices
(− △ −)

Table 12: Real data analysis for the balanced case: Statistics of model comparison

	DIC	EAIC	EBIC
ARH(1)	538329.40	557590.75	798437.35
ARMAH(1,1)	531329.84	553644.24	832666.56
AD	**496717.23**	**499933.89**	**540155.45**
HT	497262.67	501614.72	556033.37

All the structured covariance models were compared through DIC, EAIC and EBIC. The results are presented in Table 12. They indicate that the AD model outperforms the others. Therefore, we will continue the analysis with the selected model.

Figure 4 presents the observed and predicted raw scores with 95% equi-tailed credibility intervals for the six time-points. We can see that all observed scores distribution are well within the intervals, indicating that the model is well fitted to the data. Figure 5 presents the smoothed histograms of the latent trait estimates and the respective theoretical curves, i.e., the densities of the normal distributions considering the population parameter estimates. In general the curves are close, except for grades 4 (1999/Apr), 6 and 8, which indicates that a skewed distributions, as in Santos et al. (2013), may provide a better. Figure 6 presents Bayesian p-values based on the statistic given by Equation (37) for each item. Items with p-value below .05 or above .90 were considered to be not well fitted. We can see that only eight items were not fitted properly. The misfit could be due to DIF (*Differential item functioning*) or misspecification of the item response function. For example, items 22 and 28 identified by the Bayesian p-values as not well fitted, appear in three tests (tests 1, 2 and 3). Then, the functioning of these items could be altered by the successive applications.

Table 13 presents the estimates of the population parameters along with the associate posterior standard errors and 95% equi-tailed credibility intervals. The mean and variance of the first time-point were fixed to zero and one, respectively, in order to identify the model. The estimates indicate that the population means increased from the first to the fifth time-point with a slight decreasing from the fifth to the sixth. It indicates a reasonable improvement on the knowledged
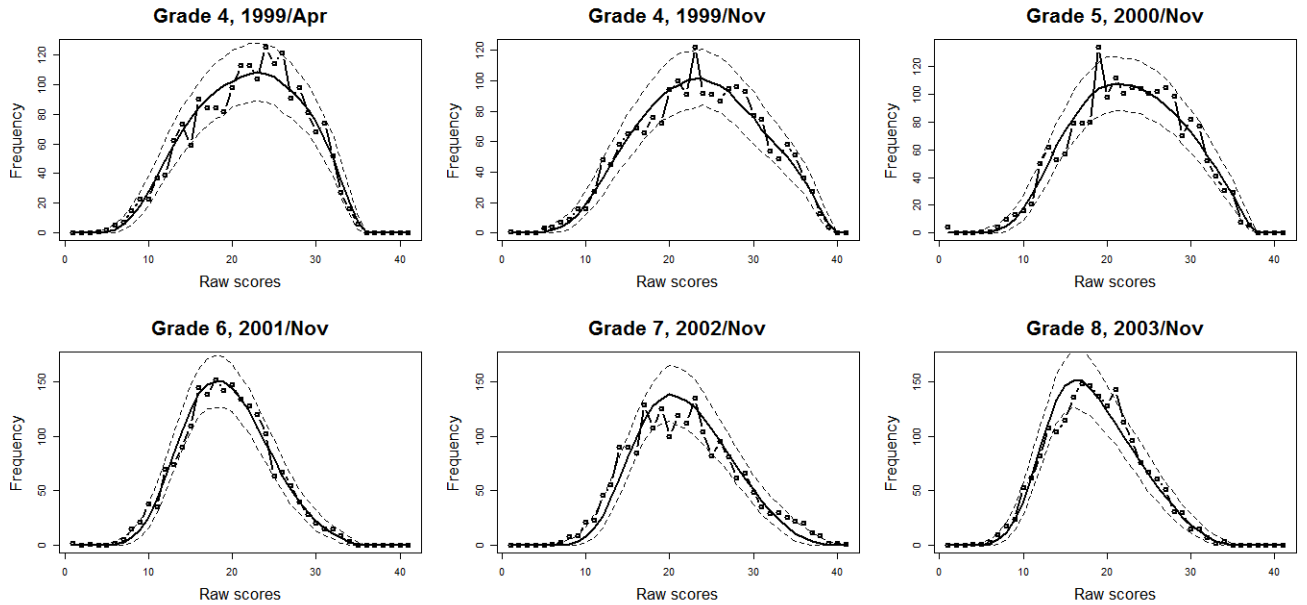
Figure 4: Real data analysis for the balanced case: observed score distribution, predicted score distribution, and 95% equi-tailed credibility intervals

in Math of the students along the first five years of the study. Also, the estimated population variances decrease from the first to the fifth time-point and increase after that, indicating that the students became more homogeneous, until the fifth test. Finally, we can observe that all correlation are significantly positive.

Figures 7 to 9 present estimates of the items parameters with their respective 95% equi-tailed credibility intervals. All tests, in general, present a reasonable discrimination power (estimates greater than .6). The discrimination power increases along the tests, except for the last one, when it decreases. It may have been caused by some failure in the tests design. Difficulty parameters tend to be higher than the mean of the latent traits along the time-points, indicating that tests are difficult. The guessing parameters estimates indicate that the actual values are different from zero, which supports the use of the three parameters model.

5.3.1 Unbalanced data

Here some observations were randomly removed from the real data, in order to generate an unbalanced structure. Therefore, the number of subjects per time are now: $n_1 = 1987$, $n_2 = 1937$, $n_3 = 1937$, $n_4 = 1917$, $n_5 = 1917$ and $n_6 = 1917$. The AD model (the choosen model in the balanced case) were fitted again and the results can be seen below. Figures 10 and 11 present the predicted scores distribution and the estimated latent trait distributions (both theoretical curve and the smoothed histograms as before), respectively. We can see that the grades six, seven and eight, at this time, display local modes. This behavior was very likely due the exclusion of the subjects. The model fit is a bit worse compared with the balanced case. Less items were found to be not well fitted, which may indicates that the dropouts may have some impact on the model fit. The posterior standard deviations and the length of the 95% equi-tailed credibility intervals are higher compared with those related to the balanced case. However, the posterior means are similar to those obtained in the balanced case, except for the population variances and some population correlations. All these behaviors are expected, since we have less information to estimate the same quantity of parameters and also do to the complex structure induced by the dropouts. Within this framework, to the best of our knowledge, there are no works that deal with unbalanced LIRT data.
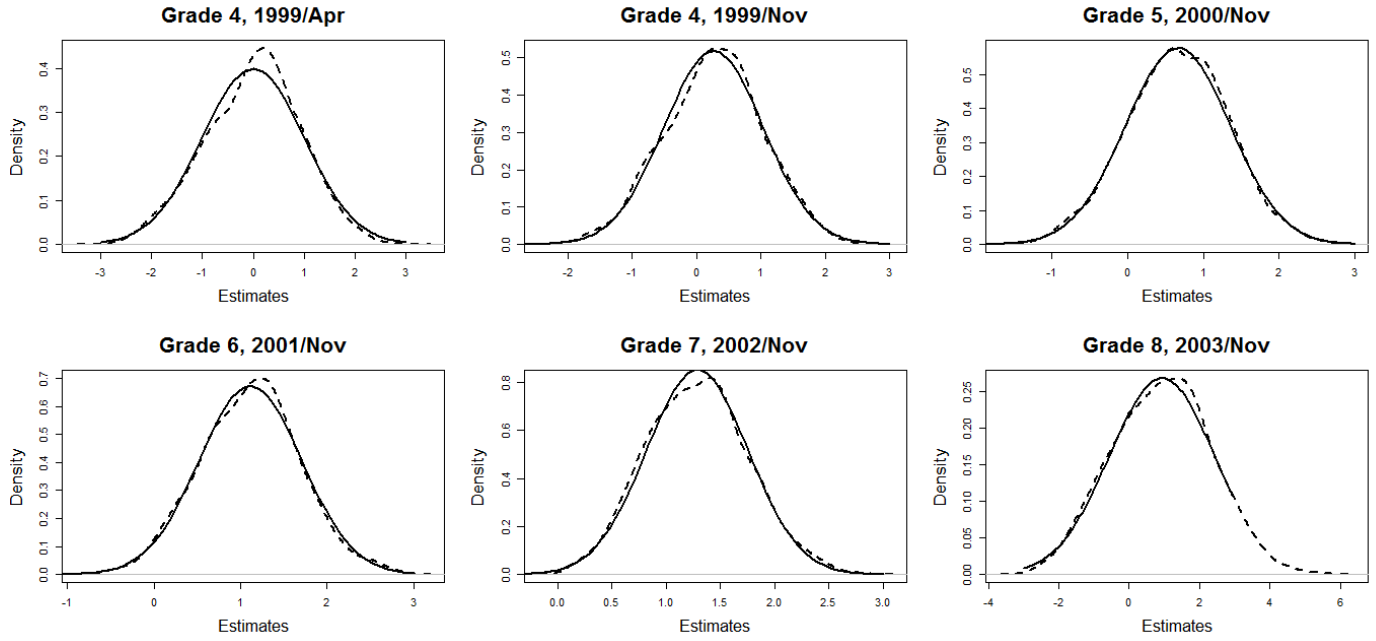
24

Figure 5: Real data analysis for the balanced case: estimated latent traits distributions. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)

Table 13: Real data analysis for the balanced case: estimates of the population parameters according to the AD model

	Mean	SD	CI (95%)		Mean	SD	CI (95%)
μ_{θ_1}	.000	–	–	ρ_{θ_1}	.784	.037	[.717, .810]
μ_{θ_2}	.268	.026	[.222, .320]	ρ_{θ_2}	.702	.041	[.628, .733]
μ_{θ_3}	.665	.030	[.604, .721]	ρ_{θ_3}	.658	.035	[.595, .697]
μ_{θ_4}	1.115	.052	[.997, 1.210]	ρ_{θ_4}	.618	.022	[.580, .654]
μ_{θ_5}	1.290	.065	[1.133, 1.403]	ρ_{θ_5}	.546	.034	[.485, .584]
μ_{θ_6}	.935	.074	[.784, 1.073]				
$\sigma^2_{\theta_1}$	1.000	–	–				
$\sigma^2_{\theta_2}$	.597	.057	[.483, .715]				
$\sigma^2_{\theta_3}$	.480	.074	[.321, .667]				
$\sigma^2_{\theta_4}$	.354	.057	[.242, .472]				
$\sigma^2_{\theta_5}$	.219	.032	[.158, .287]				
$\sigma^2_{\theta_6}$	2.255	.610	[1.008, 3.588]				

Figure 6: Real data analysis for the balanced case: Bayesian p-values for item parameters
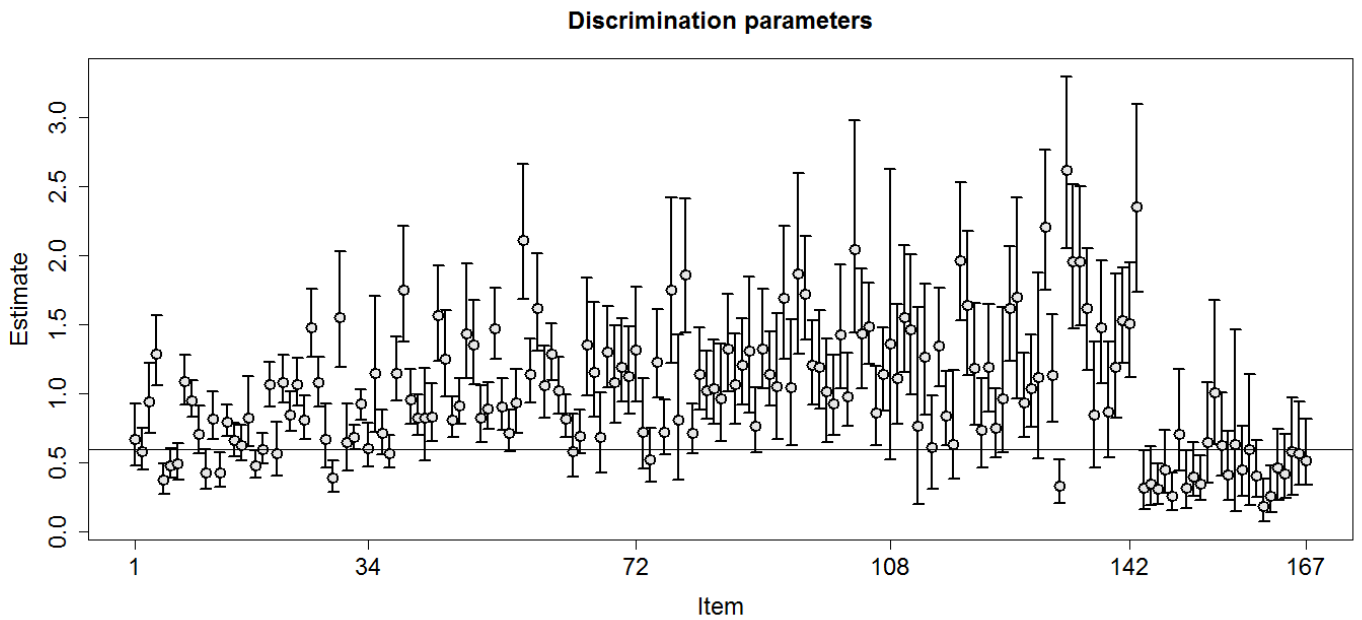
Figure 7: Real data analysis for the balanced case: posterior means and 95% equi-tailed credibility intervals for discrimination parameters
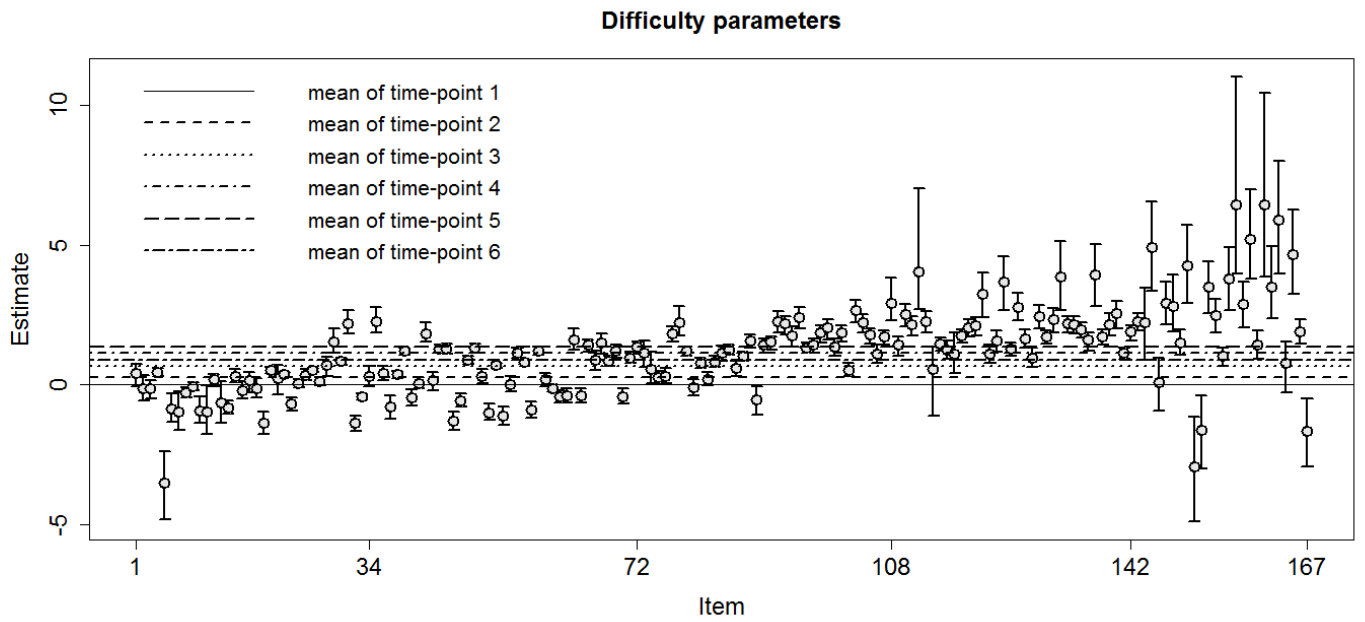
Figure 8: Real data analysis for the balanced case: posterior means and 95% equi-tailed credibility intervals for difficulty parameters
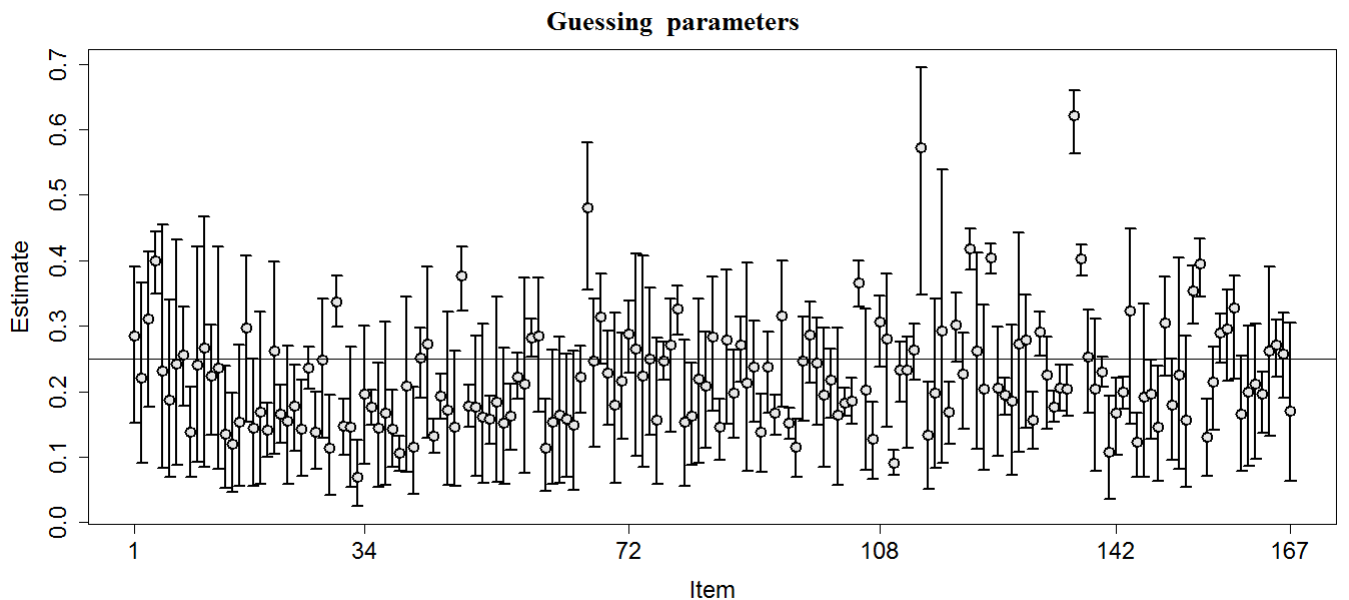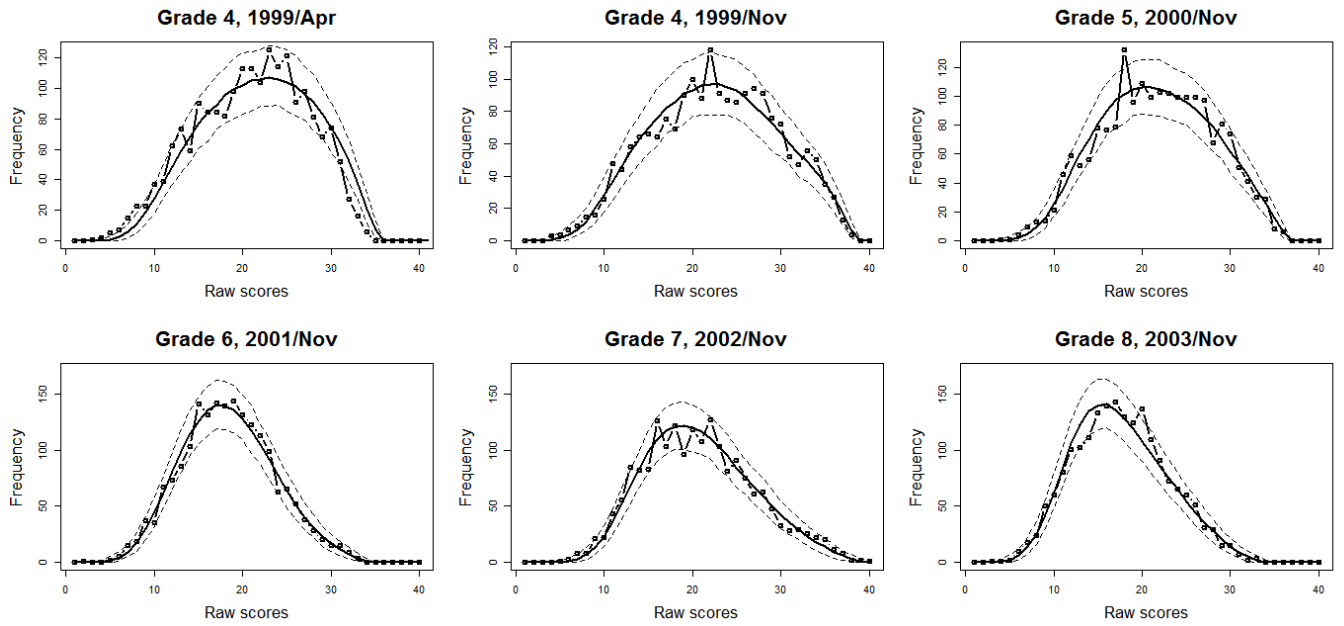

Figure 9: Real data analysis for the balanced case: posterior means and 95% equi-tailed credibility intervals for guessing parameters

Figure 10: Real data analysis for the unbalanced case: observed and predicted scores distributions with 95% equi-tailed credibility intervals

Table 14: Real data analysis for the unbalanced case: estimates of the population parameters according to the AD model

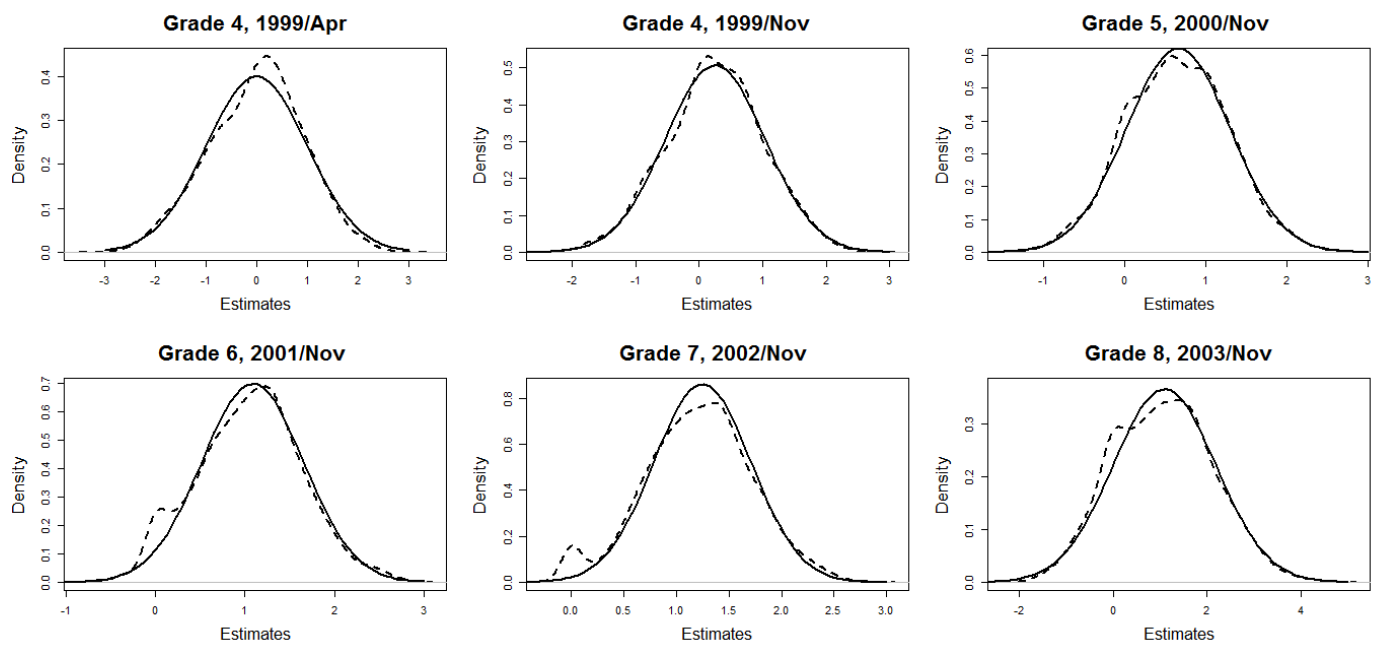|	Mean	SD	CI (95%)		Mean	SD	CI (95%)
μ_{θ_1}	.000	–	–	ρ_{θ_1}	.785	.002	[.784, .793]
μ_{θ_2}	.253	.027	[.200, .308]	ρ_{θ_2}	.701	.002	[.700, .712]
μ_{θ_3}	.658	.027	[.603, .710]	ρ_{θ_3}	.652	.002	[.652, .662]
μ_{θ_4}	1.087	.037	[1.016, 1.159]	ρ_{θ_4}	.599	.001	[.594, .599]
μ_{θ_5}	1.248	.045	[1.167, 1.335]	ρ_{θ_5}	.520	.005	[.495, .521]
μ_{θ_6}	1.095	.064	[.955, 1.208]				
$\sigma^2_{\theta_1}$	1.000	–	–				
$\sigma^2_{\theta_2}$	.621	.045	[.535, .710]				
$\sigma^2_{\theta_3}$	.413	.037	[.343, .486]				
$\sigma^2_{\theta_4}$	.325	.031	[.268, .385]				
$\sigma^2_{\theta_5}$	.215	.020	[.179, .257]				
$\sigma^2_{\theta_6}$	1.200	.228	[.870, 1.868]				

Figure 11: Real data analysis for the unbalanced case: estimated latent traits distributions. Legend: Smoothed histograms (dashed line), Theoretical curve (Solide line)
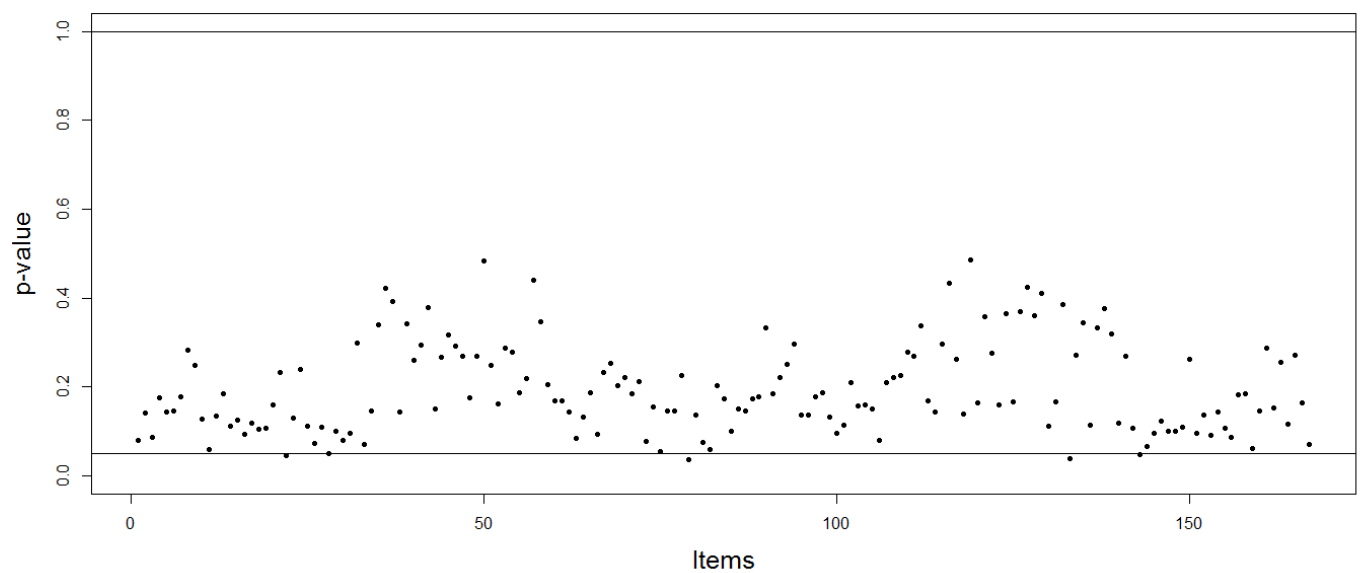

Figure 12: Real data analysis for the unbalanced case: Bayesian p-values for item parameters.
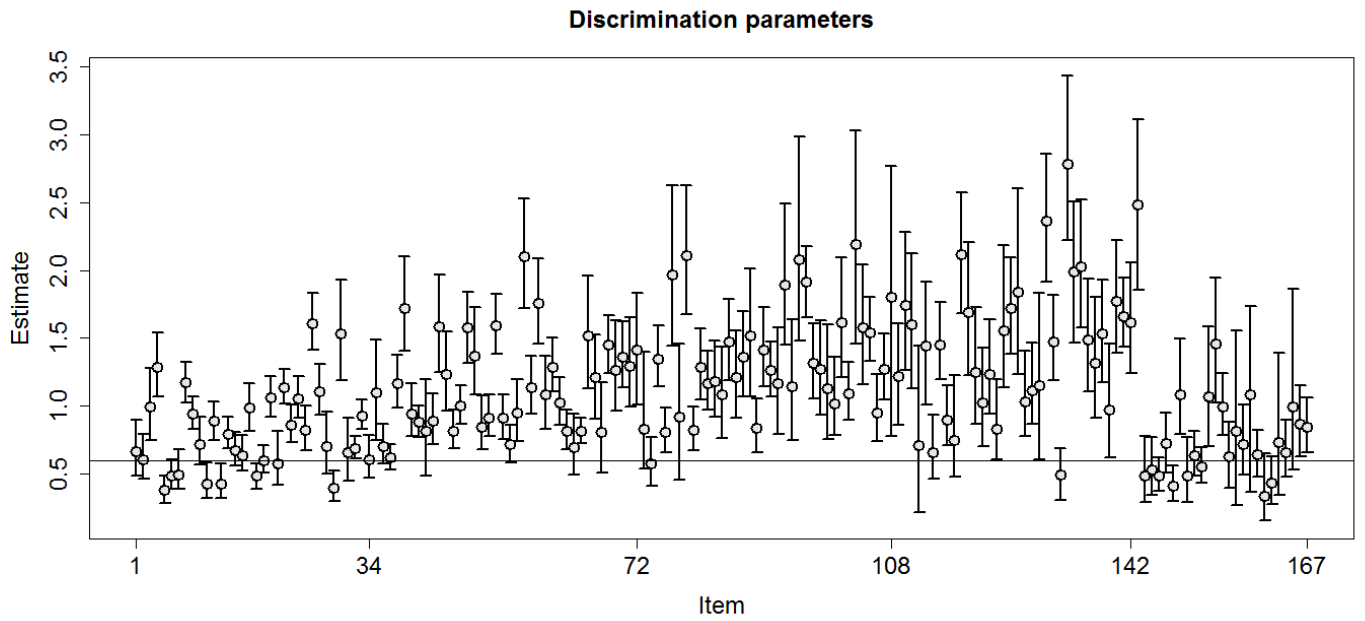
Figure 13: Real data analysis for the unbalanced case: posterior means and 95% equi-tailed credibility intervals discrimination parameters (Unbalanced case).
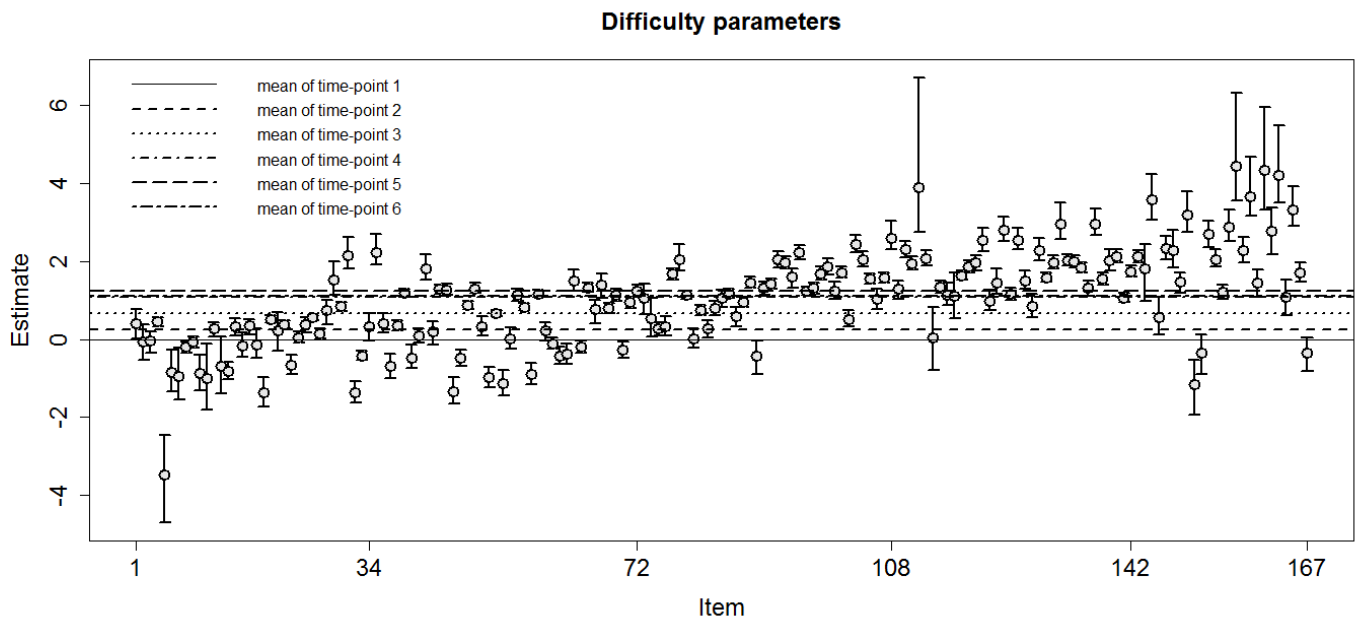

Figure 14: Real data analysis for the unbalanced case: posterior means and 95% equi-tailed credibility intervals difficulty parameters (Unbalanced case).
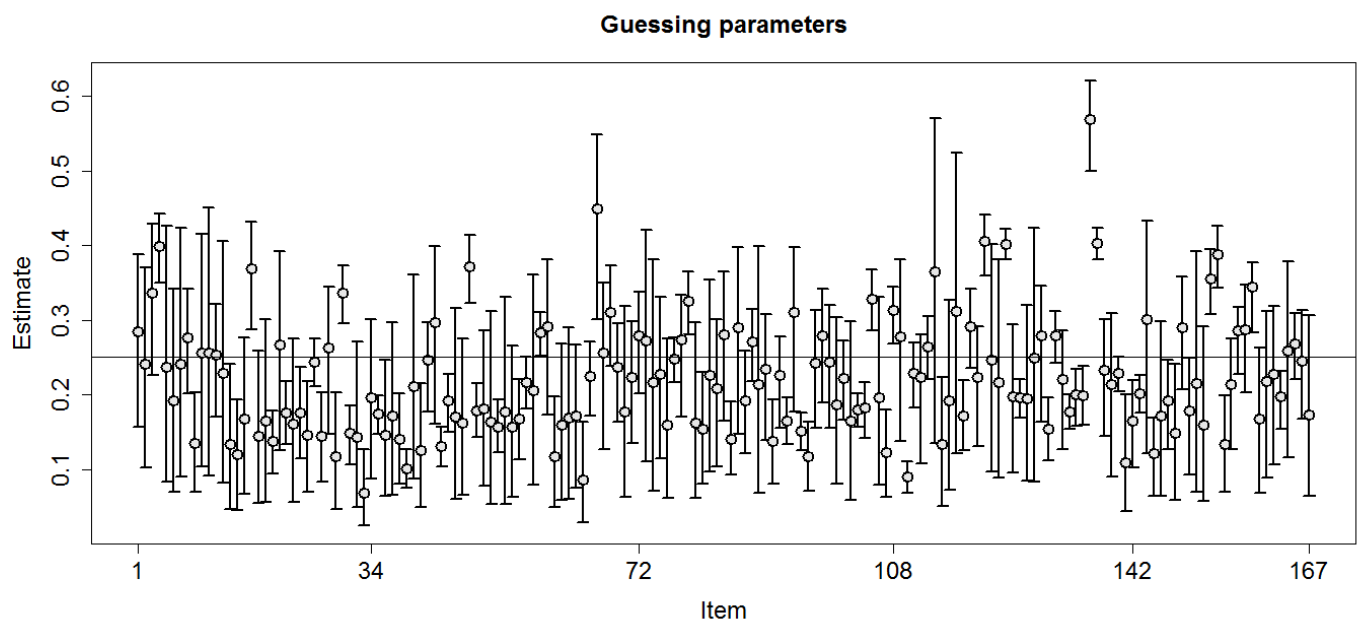
Figure 15: Real data analysis for the unbalanced case: posterior means and 95% equi-tailed credibility intervals guessing parameters (Unbalanced case).

6 Concluding Remarks

We presented a longitudinal three parameter IRT model based on a general Cholesky decomposition framework of the latent trait covariance structure. Such model is very flexible, in the sense that, accommodates properly a wide range of specific correlation patterns. The induced univariate conditional structure of the latent traits showed to be very useful to handle unbalanced data, concerning the subjects. An MCMC algorithm, based on the FFBS and SVE algorithms, was developed for parameter estimation and model fit assessment. The FFBS presented a better performance, in terms of convergence, compared to the usual approach of considering univariate full conditional distributions for the latent traits simulation. The SVE algorithm with oversampling showed to be a general approach to simulate correlation parameters. Its implementation is relatively easy for many correlation patterns by specifying suitable sufficient statistics for the correlation parameters and its corresponding theoretical autocorrelation function. This algorithm also reduces the number of parameters to be estimated, compared to the unstructured model and the multivariate approaches of Azevedo et al. (2016) and Azevedo et al. (2015), and this number does not depended on the time-points. Furthermore, our MCMC algorithm showed to be efficient in terms of parameter recovery, according to the simulation study. We performed a real data analysis using four different structured covariance matrices. The antedependence matrix fitted better to the data. As expected, the estimates under unbalanced data are less accurate than those obtained under balanced data. In a general way, our model fitted the real data well and described properly the correlation pattern of the latent traits. In conclusion, our approach revealed to be a promising alternative to the usual ones in analyzing longitudinal IRT data, speciallt unbalanced data. In future research we intend to explore some extensions of our model in order to consider other latent trait distributions, growth curves and regression structures for the mean of the latent trait distributions.

References

Andrade, D. and Tavares, H. (2005). Item response theory for longitudinal data: population parameter estimation. *Journal of Multivariate Analysis*, 95(1):1–22. 2

Azevedo, C. (2008). *Longitudinal multiple-group multilevel IRT models: estimation methods and structural selection under a Bayesian framework (PhD thesis in portugues)*. IME-USP, São Paulo, SP. 2

Azevedo, C. L., Andrade, D. F., and Fox, J.-P. (2012). A bayesian generalized multiple group irt model with model-fit assessment tools. *Computational Statistics & Data Analysis*, 56(12):4399 – 4412. 2, 15

Azevedo, C. L., Fox, J.-P., and Andrade, D. F. (2015). Longitudinal multiple-group irt modelling: covariance pattern selection using mcmc and rjmcmc. *International Journal of Quantitative Research in Education*, 2(3-4):213–243. 32

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, 4, 8, 22, 32

Baker, F. B. and Kim, S.-H. (2004). *Item response theory : Parameter Estimation Techniques*. Marcel Dekker, Inc, New York, NY, second edition. 2, 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, 21

Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an em algorithm. *Psychometrika*, 46(4):433–459. 2

Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81:541–553. 10

Conaway, M. R. (1990). A random effects model for binary data. *Biometrics*, 46(2):317–328. 2

Eid, M. (1996). Longitudinal confirmatory factor analysis for polytomous item responses : Model definition and model selection on the basis of stochastic measurement theory. *Methods of Psychological Research Online*, 1(4):65–85. 2

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis*, 15:183–202. 10

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo : Stochastic simulation for bayesian inference*. Chapman & Hall/CRC, London, second edition. 10, 11, 12

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42(4):805–820. 4

Marsman, M. (2014). *Plausible Values in Statistical Inference*. PhD Thesis University of Twente, Enschede., 1 edition. 12

Montgomery, D. C. (2004). *Design and Analysis of Experiments*. Chapman & Hall, London, sixth edition. 7

Murray, I., Ghahramani, Z., and MacKay, D. (2012). Mcmc for doubly-intractable distributions. ArXix e-prints. 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, 4

Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. *Markov Chain Monte Carlo in Pratice*, pages 45–57. 13

Sahu, S. K. (2002). Bayesian estimation and model choice in item response models. *Journal of Statistical Computation and Simulation*, 72(3):217–232. 13

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. 15, 23

Sinharay, S. (2006). A bayesian item fit analysis for unidimensional item response theory models. *British journal of mathematical and statistical psychology*, 59:429–449. 21

Sinharay, S., Johnson, M. S., and Stern, H. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, 30(4):298–321. 21

Spiegelhaltere, 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. 21

Tavares, H. R. and Andrade, D. F. (2006). Item response theory for longitudinal data: Item and population ability parameters estimation. *TEST*, 15(1):97–123. 2, 8

Zimmerman, D. and Núñez-Antón, V. (2009). *Antedependence Models for Longitudinal Data.* Chapman and Hall/CRC, 1 edition. 4