

# A Copula Based Modeling for Longitudinal IRT Data with skewed latent distributions.

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

## Abstract

In this work we introduce longitudinal IRT model considering skewed latent traits distribution, based on a Gaussian copula function. Differently of the antedependence approach proposed by Santos et al. (2017a) and Santos et al. (2017b), the copula modeling allows the entire control of the respective marginal latent trait distributions, but as the first one, it accommodates several dependence structures. 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. Estimation algorithms, model fit assessment and model comparison tools were developed under the Bayesian paradigm through hybrid MCMC algorithms, 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 in education, promoted by the Brazilian federal government, is analyzed to illustrate our methodology.

**keywords:** longitudinal IRT data, Bayesian inference, antedependence models, MCMC algorithms, Copula modeling.

---

\*Department of Statistics, University of Campinas, Brazil

†Department of Statistics, University of Campinas, Brazil

# 1 Introduction

Copula models have become one of the most widely used tools in modeling multivariate data. They have been extensively applied in survival analysis, see [Clayton \(1978\)](#) and [Oakes \(1989\)](#), actuarial science ([Frees and Valdez, 1998](#)), finance ([Li, 1999](#); [Cherubini et al., 2004](#)), marketing ([Danaher and Smith, 2011](#)), among other fields. Copulas are popular because they are flexible tools for modeling complex relationships among variables in a simple manner. They allow first modeling the marginal distributions, and then the dependence structure among the variables is captured using a suitable copula function.

Despite its great potential concerning the construction of dependence structures, there are few copula applications for modeling serial dependence in longitudinal data. Exceptions are the works of [Jiafeng et al. \(2008\)](#) and [Smith et al. \(2010\)](#). The former used copulas to model the longitudinal dependence over time and regression models with heavy-tailed distributions. The second considers a sequence of bivariate copulas, called *pair-copula* models, to model dependence structures.

We can also observe few applications of the copula approach for latent variables modeling. [Braeken et al. \(2007\)](#) highlights the potential of copula functions to handle residual dependence in IRT Rasch models. [Doebler and Doebler \(2012\)](#) proposed a class of compensatory multidimensional IRT models, see [Reckase \(2009\)](#). They used probit and logistic models from Rasch family, combined through copula functions.

In the present work we developed a copula based approach to analyze longitudinal IRT data with skewed marginal latent trait distributions. This modeling is an alternative to the Cholesky based approach, proposed by [Santos et al. \(2017b\)](#). A clear advantage of the copula approach with respect to the antedependence models is its great flexibility to define the marginal latent trait distributions. According to the error term distributions, the marginal latent trait distributions are not easy to obtain. In the copula modeling the marginal distributions are directly specified and they do not need to be from the same family, that is, we can define different family distributions for each marginal. Therefore, it is possible to consider, heavy-tailed and multimodal latent trait distributions, for example.

Basically, there are three kind of copula function: Archimedean ([Nelsen, 2007](#)), elliptical ([Frahm et al., 2003](#)) and D-vine ([Bedford and Cooke, 2002](#)). In this work, we will use Gaussian copula. This is an elliptical copula that allows to define specific covariance matrices to handle dependence pattern of the latent traits (see the next section for more

details).

This paper is outlined as follows. In section 2 we introduce the model and present their main properties. In section 3 we describe the MCMC algorithm developed to parameter estimation. In section 4 a comparison between the copula and antedependence approaches are performed by using simulated data. In section 5 a real data set is analyzed and the copula and antedependence models are compared. Finally, in section 6 we presented some comments and conclusions.

## 2 Modeling

A copula function can be defined as a  $T$ -dimensional distribution function  $C : [0, 1]^T \rightarrow [0, 1]$ , such that all its marginal distributions are uniformly distributed on  $[0, 1]$ . Other definitions can be find in Nelsen (2007). The Sklar's theorem (Skalar, 1959) ensures that for any  $T$ -dimensional distribution function  $F_{\mathbf{X}}$  with univariate marginal  $F_{X_1}, F_{X_2}, \dots, F_{X_T}$  there exist a copula function  $C$  such that

$$F_{\mathbf{X}}(x_1, \dots, x_T) = C(F_{X_1}(x_1), \dots, F_{X_T}(x_T)). \quad (1)$$

If  $F_{X_1}, F_{X_2}, \dots, F_{X_T}$  are all continuous, so  $C(\cdot)$  is unique. We further assume that  $C$  is differentiable. If we let  $c$  denote the density associated with copula  $C$ , then the density  $f$  of the  $T$ -dimensional distribution  $F$  is given by

$$f(x_1, \dots, x_T) = c(F_1(x_1), \dots, F_T(x_T)) \prod_{t=1}^T f_t(x_t). \quad (2)$$

These results are extremely important do build the skew longitudinal IRT model via copula, as we present below.

### 2.0.1 Skew longitudinal IRT model via copulas

The IRT data structure consists in  $T$  time-points which one with  $n_t$  subjects ( $t = 1, 2, \dots, T$ ), 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  $n = \sum_{t=1}^T n_t$ . Dropouts and inclusions of subjects during the study are allowed. Let us define the following notation:  $\theta_{jt}$  is the latent trait of the subject  $j$  ( $j = 1, 2, \dots, n_t$ ) at the time-point  $t$ ,  $\boldsymbol{\theta}_j$  is the latent traits vector of the subject  $j$ , and  $\boldsymbol{\theta}_{..}$  is the vector of all latent traits. Let

$Y_{ijt}$  denoting the response of the subject  $j$  to the item  $i$  ( $i = 1, 2, \dots, I$ ) at the time-point  $t$ ,  $\mathbf{Y}_{\cdot jt} = (Y_{1jt}, \dots, Y_{Ijt})'$  is the response vector of subject  $j$  at the time-point  $t$ ,  $\mathbf{Y}_{\dots} = (\mathbf{Y}'_{\cdot 1}, \dots, \mathbf{Y}'_{\cdot T})'$  is the entire response matrix and  $(y_{ijt}, \mathbf{y}'_{\cdot jt}, \mathbf{y}'_{\dots})'$  are the respective observed values. Let  $\boldsymbol{\zeta}_i$  represents the vector of item parameters of the 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.

Therefore our skew IRT longitudinal 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 suitable for dichotomous or dichotomized responses. In the second, the multivariate structure of the latent traits is modeled through the copula model. That is,

$$\begin{aligned} Y_{ijt} | \theta_{jt}, \boldsymbol{\zeta}_i &\sim \text{Bernoulli}(P_{ijt}), \\ P_{ijt} = P(Y_{ijt} | \theta_{jt}, \boldsymbol{\zeta}_i) &= c_i + (1 - c_i) \Phi(a_i \theta_{jt} - b_i). \end{aligned} \quad (3)$$

$$\boldsymbol{\theta}_j | \boldsymbol{\eta}_\theta \sim G_T(\boldsymbol{\mu}_\theta, \boldsymbol{\sigma}_\theta, \boldsymbol{\gamma}_\theta, \boldsymbol{\varphi}), \quad (4)$$

where  $G_T$  denotes a general  $T$ -dimensional distribution, whose margins are centered skew-normal distributions, see [Azzalani \(1985\)](#) and [Pewsey \(2000\)](#), denoted by  $CSN(\mu_{\theta_t}, \sigma_{\theta_t}^2, \gamma_{\theta_t})$ , where  $(\mu_{\theta_t}, \sigma_{\theta_t}^2, \gamma_{\theta_t})$  denote the population mean, variance and Pearson's skewness coefficient at time point  $t$ , following notation of [Santos et al. \(2017b\)](#).

Thus the joint density is given by

$$g(\theta_{j1}, \dots, \theta_{jT}; \boldsymbol{\varphi}) = c(F_{\theta_{j1}}, \dots, F_{\theta_{jT}}; \boldsymbol{\varphi}) \prod_{t=1}^T f_{\theta_{jt}}, \quad (5)$$

where  $F_{\theta_{jt}}$  e  $f_{\theta_{jt}}$  denotes the distribution function and the margins for latent trait of the subject  $j$  at time-point  $t$ , respectively. The vector  $\boldsymbol{\varphi}$  stands for some set of parameters associated to the copula  $C$ .

## 2.0.2 Choice of copulas

An important issue for development of the copula approach, is the selection of the most appropriate one for modeling the dependence structure of the data in study. Archimedean

copulas (Nelsen, 2007) and elliptical copulas (Frahm et al., 2003), are widely used in the literature. Archimedean copulas present good mathematical properties and are easily constructed through an unique convex function. Among the most used are the Frank, Gumbel and Clayton ones, see Nelsen (2007) and Joe (1997) for more details. Elliptical copulas, in their turn, are extracted from elliptical distributions. Two important cases are the Gaussian and  $t$ -Student copulas.

For longitudinal data applications, elliptical copulas are more useful than Archimedean ones. In longitudinal data, the dependence among the within-subjects observations is typically a function of time. For example, the dependence may decrease exponentially with the increasing between the distance of the measurement occasions. Elliptical copulas present correlation structures that can handle the serial correlation typically presented in longitudinal data. On the other hand, Archimedean copulas are symmetric among observations over time. In this sense, Archimedean copulas are not appropriate for describing the serial dependence of longitudinal data, especially when the number of time-points (many measurement conditions) is large. However, Archimedean copulas may be useful in longitudinal data applications with small number of time-points.

Another alternative when the observations are time-ordered is to consider the so-called *D-vine* copulas. These kind of copulas are constructed from sequences of bivariate copulas. Considering this construction, any mix of bivariate copulas (called *pair-copulas*) can be used, resulting in an extremely flexible modeling framework. Furthermore, the *D-vine* approach fully exploits the time ordering of the margins, and it may lead to more accurate estimates, see Smith et al. (2010). In this work we will consider Gaussian copulas.

### 2.0.3 The Gaussian copula model

The Gaussian copula (Xue-Kun Song, 2000) is the most popular of the elliptical copulas. It is defined from the multivariate normal distribution. Let  $\Phi_T(\cdot; \mathbf{R}_\theta)$  be the distribution function of a  $T$ -dimensional normal distribution with zero mean and correlation matrix  $\mathbf{R}_\theta$  and  $\Phi(\cdot)$  be the distribution function of a unidimensional standard normal distribution. The Gaussian copula function for a subject  $j$  is defined by

$$C(u_{j1}, \dots, u_{jT}; \mathbf{R}_\theta) = \Phi_T(\Phi^{-1}(u_{j1}), \dots, \Phi^{-1}(u_{jT}); \mathbf{R}_\theta). \quad (6)$$

Derivation of the copula density is straightforward by differentiation of the equation (6), so that

$$c(u_{j1}, \dots, u_{jT}; \boldsymbol{\eta}_\theta, \mathbf{R}_\theta) = \frac{\partial}{\partial \mathbf{u}_j} C(\mathbf{u}_j) = |\mathbf{R}_\theta|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}'_j \mathbf{R}_\theta^{-1} \mathbf{x}_j \right\}. \quad (7)$$

Therefore, according to the equations (5) and (7) the joint probability density of the latent traits for a subject  $j$  is given by

$$p(\boldsymbol{\theta}_j | \boldsymbol{\eta}_\theta, \mathbf{R}_\theta) = |\mathbf{R}_\theta|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}'_j \mathbf{R}_\theta^{-1} \mathbf{x}_j \right\} \prod_{t=1}^T f_t(\theta_{jt}; \boldsymbol{\eta}_{\theta_t}), \quad (8)$$

where  $\mathbf{x}_j = (x_{j1}, \dots, x_{jT})'$ ,  $x_{jt} = \Phi^{-1}(u_{jt})$ ,  $u_{jt} = F_j(\theta_{jt}; \boldsymbol{\eta}_{\theta_t})$  and matrix  $\mathbf{R}_\theta$  will be assumed to be a structured correlation matrix. Table 1 presents four possible structures of  $\mathbf{R}_\theta$ . The marginal densities and marginal distributions functions of the latent traits are denoted by  $f_t(\theta_{jt}; \boldsymbol{\eta}_{\theta_t})$  and  $F_t(\theta_{jt}; \boldsymbol{\eta}_{\theta_t})$ , respectively and they are assumed to be centered skew-normal distribution with density given by

$$f_t(\theta_{jt} | \boldsymbol{\eta}_{\theta_t}) = 2\omega_t^{-1} \phi(\omega_t^{-1}(\theta_{jt} - \xi_t)) \Phi[\lambda_t(\omega_t^{-1}(\theta_{jt} - \xi_t))], \quad (9)$$

which correspond to a usual skew-normal distribution with parameters defined as:

$$\begin{aligned} \xi_t &= \mu_{\theta_t} - \sigma_{\theta_t} \gamma_{\theta_t}^{1/3} s, \\ \omega_t &= \sigma_{\theta_t} \sqrt{1 + \gamma_{\theta_t}^{2/3} s^2}, \\ \lambda_t &= \frac{\gamma_{\theta_t}^{1/3} s}{\sqrt{r^2 + s^2 \gamma_{\theta_t}^{2/3} (r^2 - 1)}}, \text{ where} \\ s &= \left(\frac{2}{4 - \pi}\right)^{1/3}. \end{aligned} \quad (10)$$

### 3 Bayesian Estimation and MCMC Algorithms

In order to facilitate the implementation of the MCMC algorithms, particularly, aiming to obtain full conditional distribution with known form and to develop proper model-fit assessment tools; we will use the augmented data approach to represent our IRT model, see [Tanner and Wong \(1987\)](#). Essentially, we intend to obtain as many as possible of full

Table 1: Structured correlation matrices used in this work.

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

conditional distributions with known form. For the three-parameters models we can use the augmented data scheme proposed by [Béguin and Glas \(2001\)](#). This methodology consist 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  $W_{ijt}$  given  $Y_{ijt} = y_{ijt}$  is given by

$$\begin{aligned}
\mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 1, \theta_{jt}, \boldsymbol{\zeta}_i) &\propto \Phi(a_i \theta_{jt} - b_i) \\
\mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 1, \theta_{jt}, \boldsymbol{\zeta}_i) &\propto c_i (1 - \Phi(a_i \theta_{jt} - b_i)) \\
\mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 0, \theta_{jt}, \boldsymbol{\zeta}_i) &= 0 \\
\mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 0, \theta_{jt}, \boldsymbol{\zeta}_i) &= 1.
\end{aligned} \tag{11}$$

Therefore the augmented variables  $\mathbf{Z} = (Z_{111}, \dots, Z_{1n_11}, \dots, Z_{In_T T})'$ , are given by

$$Z_{ijt} | (\theta_{jt}, \zeta_i, w_{ijt}) = \begin{cases} N(a_i \theta_{jt} - b_i, 1) \mathbb{I}_{(z_{ijt} \geq 0)}, & \text{if } w_{ijt} = 1, \\ N(a_i \theta_{jt} - b_i, 1) \mathbb{I}_{(z_{ijt} < 0)}, & \text{if } w_{ijt} = 0. \end{cases} \tag{12}$$

The original response can be represented by

$$Y_{ijt} = \mathbb{I}(Z_{ijt} > 0) \mathbb{I}(W_{ijt} = 1) + \mathbb{I}(Z_{ijt} \leq 0) \mathbb{I}(W_{ijt} = 0), \tag{13}$$

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

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

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

$$L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta}, |\mathbf{z}_{..}, \mathbf{w}_{..}, \mathbf{y}_{..}, \mathbf{I}) \propto \prod_{t=1}^T \prod_{j=1}^{n_t} \prod_{i \in I_{jt}} \exp \left\{ -0.5(z_{ijt} - a_i \theta_{jt} + b_i)^2 \right\} \mathbb{I}_{(z_{ijt}, w_{ijt})} \\ \times p(w_{ijt} | y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i), \quad (14)$$

where  $\mathbb{I}_{(z_{ijt}, w_{ijt})}$  stands for the indicator function  $\mathbb{I}_{(z_{ijt} < 0, w_{ijt} = 0)} + \mathbb{I}_{(z_{ijt} \geq 0, w_{ijt} = 1)}$  and  $I_{jt}$  is the set of items answered by the subject  $j$  at time  $t$  and  $p(w_{ijt} | y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i)$  follows from (11).

### 3.1 Prior and posterior distributions

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

$$p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta} | \boldsymbol{\eta}_{\zeta}, \boldsymbol{\eta}_{\eta}) = \left\{ \prod_{t=1}^T \prod_{j=1}^{n_t} p(\theta_{jt} | \boldsymbol{\eta}_{\theta_t}) \right\} \left\{ \prod_{i=1}^I p(\boldsymbol{\zeta}_i | \boldsymbol{\eta}_{\zeta}) \right\} \left\{ \prod_{t=1}^T p(\boldsymbol{\eta}_{\theta_t} | \boldsymbol{\eta}_{\eta}) \right\}, \quad (15)$$

where  $\boldsymbol{\eta}_{\zeta}$  and  $\boldsymbol{\eta}_{\eta}$  are hyperparameters associated with  $\boldsymbol{\zeta}$  and  $\boldsymbol{\eta}_{\theta}$ , respectively. Moreover, we are assuming independence between items and population parameters. The prior distributions of the latent traits are defined in equation (9). For the item parameters we have:

$$p(\boldsymbol{\zeta}_{i(-c_i)}) \propto \exp \left[ -0.5(\boldsymbol{\zeta}_{i(-c_i)} - \boldsymbol{\mu}_{\zeta})' \boldsymbol{\Psi}_{\zeta}^{-1} (\boldsymbol{\zeta}_{i(-c_i)} - \boldsymbol{\mu}_{\zeta}) \right] \mathbb{I}_{(a_i > 0)} \quad (16)$$

and

$$c_i \sim \text{Beta}(a_c, b_c), \quad (17)$$

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

$$\begin{aligned}
p(\mu_{\theta_t}, \sigma_{\theta_t}^2, \gamma_{\theta_t}, \boldsymbol{\rho}_\theta) &= p(\mu_{\theta_t})p(\sigma_{\theta_t}^2)p(\gamma_{\theta_t})p(\boldsymbol{\rho}_\theta) \\
&= p(\mu_{\theta_t})p(\sigma_{\theta_t}^2)p(\gamma_{\theta_t}) \prod_{t=1}^{T-1} p(\rho_{\theta_t})
\end{aligned} \tag{18}$$

$$\begin{aligned}
\mu_{\theta_t} &\sim N(m_\mu, \sigma_\mu^2), \\
\sigma_{\theta_t}^2 &\sim IG(a_\sigma, b_\sigma), \\
\gamma_{\theta_t} &\sim N(\mu_\gamma, \sigma_\gamma^2) \mathbb{I}_{[-.99527, .99527]}, \quad t = 1, \dots, T, \\
\rho_{\theta_t} &\sim N(\mu_\rho, \sigma_\rho^2) \mathbb{I}_{(0,1)},
\end{aligned} \tag{19}$$

where  $IG$  denotes the inverse-gamma distribution with  $\mathbb{E}(\sigma_{\theta_t}^2) = a_\sigma/b_\sigma$  and  $\text{Var}(\sigma_{\theta_t}^2) = a_\sigma/b_\sigma^2$ . The prior distribution of the correlation parameters for every considered structure are the same.

Given the augmented likelihood in equation (14) and the prior distribution in equations (8), (9), (16), (17) and (19), the joint posterior distribution is given by:

$$\begin{aligned}
p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, |\mathbf{z}_{..}, \mathbf{w}_{..}, \mathbf{y}_{..}) &\propto \left\{ \prod_{t=1}^T \prod_{j=1}^n \prod_{i \in I_{jt}} \exp \{-.5(z_{ijt} - a_i \theta_{jt} + b_i)^2\} p(w_{ijt} | y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i) \mathbb{I}_{(z_{ijt}, w_{ijt})} \right\} \\
&\times \left\{ \prod_{j=1}^{n_t} p(\boldsymbol{\theta}_j | \boldsymbol{\eta}_{\theta_t}, \mathbf{R}_\theta) \right\} \left\{ \prod_{i=1}^I p(\boldsymbol{\zeta}_i | \boldsymbol{\eta}_\zeta) \right\} \\
&\times \left\{ \prod_{t=1}^T p(\boldsymbol{\eta}_{\theta_t} | \boldsymbol{\eta}_\eta) \right\}.
\end{aligned} \tag{20}$$

Since the posterior distribution has an intractable form, we will use MCMC algorithms in order to obtain empirical approximation for the posterior marginal distributions. Due to the augmented data structure, the full conditional distributions of the item parameters are simple to sample from. For the other parameters we need to consider auxiliary algorithms. In this case we consider either the Metropolis-Hastings or the SVE, depending on the parameter. For the item parameters we have,

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

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

where  $(\cdot)$  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 the guessing parameters we consider,

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

where

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

For the other parameters, the full conditional distributions are not known and auxiliary algorithms will be needed. To simulate from the population mean, variance and skewness coefficient we consider Metropolis-Hastings steps with the following proposal densities:

$$\begin{aligned} p(\mu_{\theta_t}^{(m)} | \mu_{\theta_t}^{(m-1)}) &= N(\mu_{\theta_t}^{(m-1)}, \sigma_0^2), \\ p((\sigma_{\theta_t}^2)^{(m)} | (\sigma_{\theta_t}^2)^{(m-1)}) &= \text{Lognormal}((\sigma_{\theta_t}^2)^{(m-1)}, \sigma_0^2), \\ p(\gamma_{\theta_t}^{(m)} | \gamma_{\theta_t}^{(m-1)}) &= U\left(\nu_1(\gamma_{\theta_t}^{(m-1)}), \nu_2(\gamma_{\theta_t}^{(m-1)})\right) \text{ for all } t = 1, \dots, T, \end{aligned} \quad (23)$$

where

$$\nu_1 = \max\{-.99527, \gamma_{\gamma_t} - \Delta_\gamma\} \text{ and } \nu_2 = \max\{.99527, \gamma_{\theta_t} + \Delta_\gamma\}, \quad \Delta_\gamma > 0.$$

The constant  $\Delta_\gamma$  is previously defined, see [Azevedo et al. \(2011\)](#). The superscript  $(m)$  indicates the estimate at the iteration  $m$  of the MCMC algorithm. Algorithms [3.1](#) to [3.3](#) correspond to the Metropolis-Hastings steps to simulate population mean, variance and skewness coefficient, respectively, for all  $t = 1, \dots, T$ .

---

**Algorithm 3.1** Population mean sampler

---

- 1: Draw  $\mu_{\theta_t}^{(m)} \sim N(\mu_{\theta_t}^{(m-1)}, \sigma_0^2)$
- 2: Draw  $u \sim U(0, 1)$
- 3: **if**  $u < \min\{1, \mathcal{Q}_t\}$  where

$$\mathcal{Q}_t = \frac{\prod_{j=1}^{n_t} p(\boldsymbol{\theta}_{j.}^{(m-1)} | \boldsymbol{\eta}_\theta^{(m)}, \mathbf{R}_\theta^{(m-1)}) p(\mu_{\theta_t}^{(m)})}{\prod_{j=1}^{n_t} p(\boldsymbol{\theta}_{j.}^{(m-1)} | \boldsymbol{\eta}_\theta^{(m-1)}, \mathbf{R}_\theta^{(m-1)}) p(\mu_{\theta_t}^{(m-1)})}$$

- 4: **then**  $\mu_{\theta_t}^{(m-1)} = \mu_{\theta_t}^{(m)}$
  - 5: **end if**
- 

---

**Algorithm 3.2** Population variance sampler

---

- 1: Draw  $(\sigma_{\theta_t}^2)^{(m)} \sim \text{Lognormal}((\sigma_{\theta_t}^2)^{(m-1)}, \sigma_0^2)$
- 2: Draw  $u \sim U(0, 1)$
- 3: **if**  $u < \min\{1, \mathcal{Q}_t\}$  where

$$\mathcal{Q}_t = \frac{\prod_{j=1}^{n_t} p(\boldsymbol{\theta}_{j.}^{(m-1)} | \boldsymbol{\eta}_\theta^{(m)}, \mathbf{R}_\theta^{(m-1)}) p((\sigma_{\theta_t}^2)^{(m-1)} | (\sigma_{\theta_t}^2)^{(m-1)}) p((\sigma_{\theta_t}^2)^{(m)})}{\prod_{j=1}^{n_t} p(\boldsymbol{\theta}_{j.}^{(m-1)} | \boldsymbol{\eta}_\theta^{(m-1)}, \mathbf{R}_\theta^{(m-1)}) p((\sigma_{\theta_t}^2)^{(m-1)} | (\sigma_{\theta_t}^2)^{(m-1)}) p((\sigma_{\theta_t}^2)^{(m-1)})}$$

- 4: **then**  $(\sigma_{\theta_t}^2)^{(m-1)} = (\sigma_{\theta_t}^2)^{(m)}$
  - 5: **end if**
- 

---

**Algorithm 3.3** Population skewness coefficient sampler

---

- 1: Draw  $\gamma_{\theta_t}^{(m)} \sim U(\nu_1(\gamma_{\theta_t}^{(m-1)}), \nu_2(\gamma_{\theta_t}^{(m-1)}))$
- 2: Draw  $u \sim U(0, 1)$
- 3: **if**  $u < \min\{1, \mathcal{Q}_t\}$  where

$$\mathcal{Q}_t = \frac{\prod_{j=1}^{n_t} p(\boldsymbol{\theta}_{j.}^{(m-1)} | \boldsymbol{\eta}_\theta^{(m)}, \mathbf{R}_\theta^{(m-1)}) p(\gamma_{\theta_t}^{(m)}) [\nu_1(\gamma_{\theta_t}^{(m)}) - \nu_2(\gamma_{\theta_t}^{(m)})]}{\prod_{j=1}^{n_t} p(\boldsymbol{\theta}_{j.}^{(m-1)} | \boldsymbol{\eta}_\theta^{(m-1)}, \mathbf{R}_\theta^{(m-1)}) p(\gamma_{\theta_t}^{(m-1)}) [\nu_1(\gamma_{\theta_t}^{(m-1)}) - \nu_2(\gamma_{\theta_t}^{(m-1)})]}$$

- 4: **then**  $\gamma_{\theta_t}^{(m-1)} = \gamma_{\theta_t}^{(m)}$
  - 5: **end if**
-

### 3.2 Correlation parameters sampler

Concerning the correlation parameters, we have that the full conditional distribution of  $\rho_\theta$  is given by:

$$\begin{aligned} p(\rho_\theta | (\cdot)) &\propto |\mathbf{R}_\theta|^{-1/2} \left( \prod_{j=1}^{n_t} \exp \left\{ -\frac{1}{2} \mathbf{x}'_j \mathbf{R}_\theta^{-1} \mathbf{x}_j \right\} \right) p(\rho_\theta) \\ &= \prod_{j=1}^{n_t} p(\mathbf{x}_j | \boldsymbol{\eta}_{\theta_t}) p(\rho_\theta). \end{aligned} \quad (24)$$

We can see that random variables  $\mathbf{X}_j$  can be seen as multivariate normal random variables, that is,  $\mathbf{X}_j \sim N_T(\mathbf{0}, \mathbf{R}_\theta)$  for all  $t = 1, \dots, T$ . Therefore, an antedependence model, in the sense of that proposed by Santos et al. (2017a) can be considered. The distribution of  $\mathbf{X}_j$  can be expressed as:

$$X_{jt} = \sum_{k=1}^{t-1} \phi_{tk} X_{jk} + \varepsilon_{jt}, \quad \varepsilon_{jt} \stackrel{i.i.d.}{\sim} N(0, d_t), \quad t = 1, 2, \dots, T. \quad (25)$$

Equivalently,

$$\boldsymbol{\varepsilon}_j = \mathbf{L}_X \mathbf{X}_j, \quad \boldsymbol{\varepsilon}_j \stackrel{ind.}{\sim} N_T(\mathbf{0}, \mathbf{D}_X), \quad (26)$$

where  $\mathbf{D}_X = \text{diag}(1, d_2, \dots, d_T)$  and  $\mathbf{L}_X$  is a  $(T \times T)$  lower-triangular matrix having the following form,

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

The parameters  $\phi_{tk}$  are the so called autoregressive generalized parameters. It is possible to show that:

$$\mathbf{L}_X \mathbf{R}_\theta \mathbf{L}'_X = \mathbf{D}_X. \quad (27)$$

Note that, the parameters  $(d_2, \dots, d_T, \phi_{21}, \phi_{31}, \phi_{32}, \dots, \phi_{T(T-1)})'$  are an one-to-one mapping of the correlation parameters  $\boldsymbol{\rho}_\theta$ . This approach allow us to consider an adaptation of the SVE algorithm proposed by Santos et al. (2017a). Algorithm 3.4 describes the correlation parameter sampler considering the AR(1) matrix, for example.

---

**Algorithm 3.4** The SVE algorithm with oversampling to sample a correlation parameter considering AR(1) matrix

---

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

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

1: **for**  $q = 1$  to  $Q$  **do**

2: Draw  $\rho_{\theta_q}^{(m)} \sim p(\rho_\theta)$

3: Draw  $\mathbf{X}_q^{(m)}$  from the model (25)

4: Compute  $r_1(\mathbf{X}_q^{(m)})$  the first-order sample correlation

5: **end for**

6: Choose the  $\rho_{\theta_q}^{(m)}$  whose  $r_1(\mathbf{X}_q^{(m)})$  is closest to  $r_1(\mathbf{X}_q^{(m-1)})$

7: Set  $\rho_\theta^{(m)} = \rho_{\theta_q}^{(m)}$  and  $\mathbf{X}^{(m)} = \mathbf{X}_q^{(m)}$  the candidate values.

8: Build the AR(1) proposed matrix  $\mathbf{R}_{\rho_\theta}^{(m)}$  using `AR1.matrix()`

9: Perform the Cholesky decomposition of  $\mathbf{R}_{\rho_\theta}^{(m)}$  to obtain the matrices  $\mathbf{L}_X^{(m)}$  and  $\mathbf{D}_X^{(m)}$

10: Draw  $u \sim U(0, 1)$

11: **if**

$$\left( u < \min \left\{ 1, \frac{p(\mathbf{x}^{(m-1)} | \rho_\theta^{(m)}) p(\mathbf{x}^{(m)} | \rho_\theta^{(m-1)})}{p(\mathbf{x}^{(m-1)} | \rho_\theta^{(m-1)}) p(\mathbf{x}^{(m)} | \rho_\theta^{(m)})} \right\} \right)$$

**then**

11:  $\rho_\theta^{(m-1)} = \rho_\theta^{(m)}$

10: **end if**

---

Notation  $p(\mathbf{x} | \rho_\theta)$  represents the likelihood generated by the model (25). The function `AR1.matrix()` is based on the autocorrelation function of the AR(1) process. For structures with more than one correlation parameter, the Algorithm 3.4 can be applied for each parameter independently, see Santos (2016) for more details.

### 3.3 Latent trait parameters estimation

To sample latent trait parameters we also consider the SVE algorithm. In this case, the original likelihood of the three parameters probit model is considered, instead of augmented one. The SVE algorithm to simulate latent traits consists in simulate a number of *i.i.d* proposal values of  $\theta_j$ , each with its own response pattern, and choose those whose test score  $s(\mathbf{y}) = \sum_{t=1}^T \sum_{i \in \mathcal{I}_{jt}} y_{ijt}$  is the closest to the observed test score. Algorithm 3.5 describes the latent traits sampler.

---

**Algorithm 3.5** The SVE algorithm with oversampling to sample latent traits

---

```

1: for  $q = 1$  to  $Q$  do
2:   Draw  $\theta_{j,q}^{(m)}$  based on definition (6)
3:   Draw  $Y_q^{(m)}$  from the model (3)
4:   Compute  $s(Y_q^{(m)})$  the test score
5: end for
6: Choose the  $\theta_{j,q}^{(m)}$  whose  $s(Y_q^{(m)})$  is closest to  $s(Y)$ 
7: Set  $\theta_j^{(m)} = \theta_{j,q}^{(m)}$  and  $Y^{(m)} = Y_q^{(m)}$ 
8: Draw  $u \sim U(0, 1)$ 
9: if
      
$$\left( u < \min \left\{ 1, \frac{p(Y|\theta_j^{(m)}, \zeta^{(m-1)})p(Y^{(m)}|\theta_j^{(m-1)}, \zeta^{(m-1)})}{p(Y|\theta_j^{(m-1)}, \zeta^{(m-1)})p(Y^{(m)}|\theta_j^{(m)}, \zeta^{(m-1)})} \right\} \right)$$

then
10:   $\theta_j^{(m-1)} = \theta_j^{(m)}$ 
11: end if

```

---

Notation  $p(Y|\theta_j)$  denotes the likelihood generated by the model (3). That is,

$$p(Y|\theta_j, \zeta) = \prod_{t=1}^T \prod_{j=1}^{n_t} \prod_{i \in \mathcal{I}_{jt}} P_{ijt}^{y_{ijt}} (1 - P_{ijt})^{1-y_{ijt}}, \quad (28)$$

where  $P_{ijt} = c_i + (1 - c_i)\Phi(a_i\theta_{jt} - b_i)$ . In summary, a general algorithm to estimate the parameter's model it is combination of Metropolis-Hastings, Gibbs sampling and SVE algorithms as we can see in algorithm 3.6.

## 4 Simulation Study

In this section we study the performance of our model and the proposed estimation algorithm concerning parameter recovery. Furthermore, we compare the skew copula IRT model with the Cholesky decomposition based model proposed Santos et al. (2017b). For short, we will refer these models as copula and Cholesky models, respectively. Our ob-

---

**Algorithm 3.6** Full Sampler for the skew copula IRT longitudinal model

---

- 1: Start the algorithm by choosing suitable initial values. Repeat steps 2-9.
  - 2: Simulate  $W_{ijt}$  from  $W_{ijt}|\cdot$  for all  $i = 1, \dots, I$ ,  $j = 1, \dots, n$  and  $t = 1, \dots, T$ .
  - 3: Simulate  $Z_{ijt}$  from  $W_{ijt}|\cdot$  for all  $i = 1, \dots, I$ ,  $j = 1, \dots, n$  and  $t = 1, \dots, T$ .
  - 4: Simulate  $\theta_{jt}$  using the algorithm 3.5 for all  $j = 1, \dots, n_t$  e  $t = 1, \dots, T$ .
  - 5: Simulate  $\zeta_i$  from  $\zeta_i|\cdot$  for all  $i = 1, \dots, I$ .
  - 6: Simulate  $c_i$  from  $c_i|\cdot$  for all  $i = 1, \dots, I$ .
  - 7: Simulate  $\mu_{\theta_t}$  using the algorithm 3.1 for all  $t = 1, \dots, T$ .
  - 8: Simulate  $\sigma_{\theta_t}^2$  using the algorithm 3.2 for all  $t = 1, \dots, T$ .
  - 9: Simulate  $\gamma_{\theta_t}$  using the algorithm 3.3 for all  $t = 1, \dots, T$ .
  - 10: Simulate  $\rho_{\theta_t}$  by using SVE algorithm for all  $t = 1, \dots, T - 1$ .
- 

jective is to compare the performance of the two models (in terms of parameter recovery) when the data are generated by the copula model. For simplicity and without lose generality, the AD matrix was chosen to procedure the study, since it is the most general matrix.

Responses of  $n_t = 500$  subjects, followed along  $T = 3$  time points were simulated according to the model (3). The items parameters were fixed in the following intervals:  $a_i \in [.8, 1.4]$ ,  $b_i^* \in [-2.0, 3.8]$  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 items with low, middle and high difficulty level, with respect to the mean of the latent traits. Similarly, we tried to fixe high, middle and low discrimination. Each test has 24 items with 6 anchor items between tests 1 and 2, and also 6 common items between tests 2 and 3, so that the total of items is  $I = 60$ .

The latent traits were simulated from model 8 considering:  $\boldsymbol{\mu}_\theta = (.0, 1.0, 2.0)'$ ,  $\boldsymbol{\sigma}_\theta^2 = (1.00, 1.20, .88)'$  and  $\boldsymbol{\gamma}_\theta = (.0, .8, -.5)'$  (the vector of marginal skewness coefficients). An AD correlation matrix with parameters  $\boldsymbol{\rho}_\theta = (.8, .7)'$ , was considered. We fixed increasing values for the population means on the (0, 1) scale (which correspond, respectively, to mean and variance of the latent traits in the first time-point), meaning that, the average latent traits of the respondents increased during the study. This is an expected behavior in educational longitudinal studies, for example, see Santos et al. (2013) and Azevedo et al. (2012). The values for the population variances were fixed in order to have a increasing and then a decreasing behavior. Concerning the correlation parameters, we fixed high values in order to obtain a pattern similar to that observed in the real data. The skewness coefficients were fixed in order to consider, null, high positive and high negative asymmetry, respectively.



Table 2 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. 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 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 2: The priors distributions hyperparameters

Hyperparameters					
$\mu_\zeta$	$\Psi_\zeta$	$(m_\mu; \sigma_\mu^2)$	$(a_\sigma, b_\sigma)$	$(\mu_\rho, \sigma_\rho^2)$	$(\mu_\gamma, \sigma_\gamma^2)$
(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 values of Gelman-Rubin's statistics were close to one for all parameters, indicating convergence of the MCMC algorithm. The trace plots and the values of Geweke's statistics indicate that a Burn-in of 5,000 iterations was enough to reach the convergence. Furthermore, the correlograms indicate that the samples composed by storing values at every 30th iterations have negligible autocorrelation. Therefore, we will work with valid samples with size 1,000.

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:  $\frac{1}{n_p} \sum_{l=1}^{n_p} (\hat{\vartheta}_l - \tilde{\vartheta}_l)^2$  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.

Table 3 presents the estimates of the population parameters for both models. The estimates are not so different except for the skewness coefficient. Indeed, the marginal skewness coefficient of the Cholesky model is quite influenced by the population corre-

lations, unlike the copula model. Furthermore, the copula model recovered properly all parameters. Figure 1 presents the estimated latent trait distribution with theoretical curves of the models. Figure 1 presents the true latent trait distribution along with the estimated curves of the copula and Cholesky models. We can see that, the Cholesky model describes poorly the latent trait behavior. Table 4 presents the results for the latent traits and item parameters. The Cholesky model presents higher absolute bias specially for the latent traits and item parameters. The variance related to the estimates were approximately three times higher for the latent traits, under the Cholesky model. Figures 2 and 3 present the estimates of the latent traits and item parameters with 95% credibility intervals. Some discrimination and guessing parameters are not covered by credibility intervals under the Cholesky model. Concerning the difficulty parameters the estimates under the Cholesky model estimates are similar to those obtained through the copula model, except for the 60th item whose credibility interval presented a higher length under the Cholesky model.

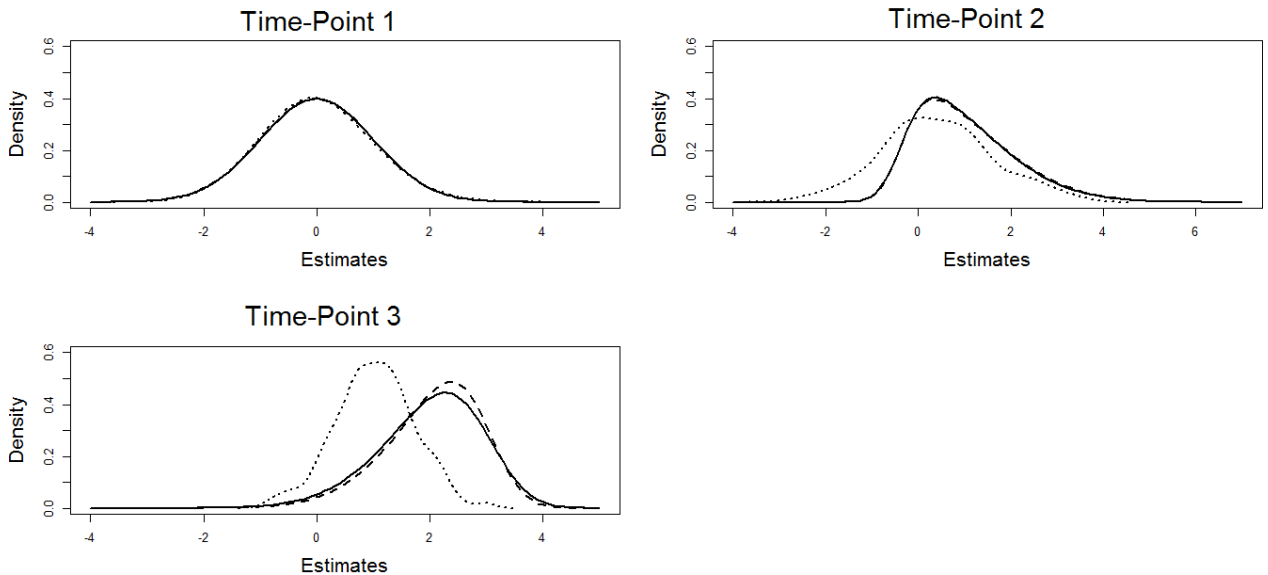


Figure 1: Estimated latent traits distributions. Legend: True distributions (solid line), Copula model (dashed line) and Cholesky model (dotted line)

Lastly, we compare the models by using DIC, EAIC and EBIC criteria, see Spiegelhalter et al. (2002). The smaller is the value of statistic the better is the model fit. The copula model was chosen by all statistics.

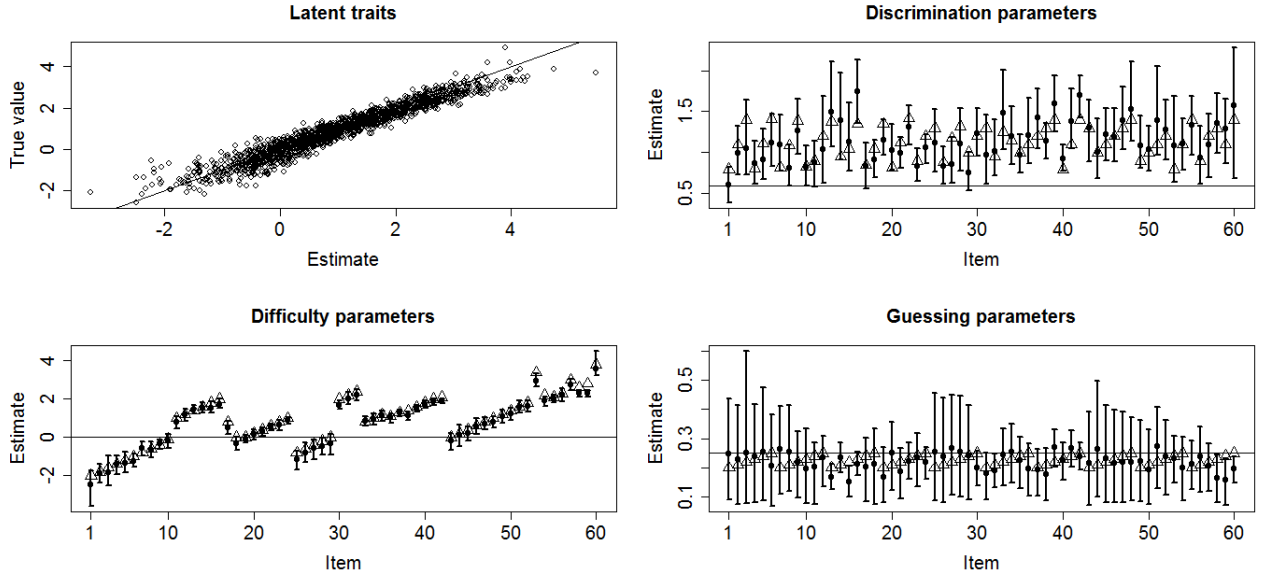


Figure 2: Estimates of latent traits and item parameters with 95% credibility intervals under the copula model. Legend: Circles denotes estimates, triangles denotes true values.

## 5 The Brazilian school development study

The analyzed data concerns to a major study initiated by the Brazilian Federal Government know as the School Development Program. It aims to monitoring 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, this is a longitudinal study, performed to evaluate children ability in Math and Portuguese language. Only the results concerning to the math part were considered in our analysis. In the present analysis we consider a subset of the data analyzed by [Santos et al. \(2017a\)](#) and [Santos et al. \(2017b\)](#). It consists in a total of 500 students, that were randomly selected, from the fourth and fifth grade of the primary school, followed in three different occasions: 1999/April, 1999/November and 2000/November with 500 students in each occasion. A total of 72 test items was used. Table 6 presents the structure of the tests, that is, the number of items per test and the number of common items across them.

In this analysis we also compare the copula model with the Cholesky decomposition based model proposed by [Santos et al. \(2017b\)](#). For short, we will refer the two models only by copula model and Cholesky model. Following the simulated study we consider the AD dependence structure in both models. Figures 4 and 5 presents the observed and predicted scores with 95% credibility intervals for both models. We can see that all

observed scores distribution are well within the intervals, indicating that both model are well fitted. Figure 6 and 7 present smoothed histograms of the latent trait estimates under copula and Cholesky models, respectively, along with theoretical curves. The theoretical curves of the copula model are defined in equation 9 and the curves of the Cholesky model were simulated via model 6. In both cases we can see that the theoretical curves are quite close to the smoothed histograms. A comparison between the estimates of the population parameters of the copula and Cholesky models can be seen in Table 7. The estimates of the correlation parameters are close, differently from those related to the other parameters. The Cholesky model presents smaller standard errors for the most of the population parameters estimates which not, necessarily, indicates that the Cholesky model is better fitted to the data.

Table 3: Results for the estimated population parameters

	Copula				Cholesky		
	True value	Mean	SD	CI (95%)	Mean	SD	CI (95%)
$\mu_{\theta_1}$	.000	.000	–	–	.000	–	–
$\mu_{\theta_2}$	1.000	1.031	.070	[.889, 1.176]	.977	.068	[.841, 1.110]
$\mu_{\theta_3}$	2.000	2.058	.080	[1.890, 2.222]	2.140	.078	[1.978, 2.271]
$\sigma_{\theta_1}^2$	1.000	1.000	–	–	1.000	–	–
$\sigma_{\theta_2}^2$	1.200	1.257	.122	[1.046, 1.516]	1.786	.286	[1.377, 2.443]
$\sigma_{\theta_3}^2$	.880	.754	.060	[.648, .872]	.561	.090	[.395, .741]
$\rho_{\theta_1}$	.800	.806	.011	[.783, .828]	.804	.015	[.776, .834]
$\rho_{\theta_2}$	.700	.701	.031	[.640, .758]	.751	.025	[.699, .802]
$\gamma_{\theta_1}$	.000	.028	.101	[-.160, .205]	.146	.136	[-.013, .457]
$\gamma_{\theta_2}$	.800	.822	.091	[.606, .939]	.136	.084	[.012, .330]
$\gamma_{\theta_3}$	-.500	-.594	.097	[-.745, -.396]	.078	.048	[.007, .210]

Table 4: Results for the estimated latent traits and item parameters.

	Parameter	Statistic				
		Corr	MBias	MABias	MVAR	MRMSE
Copula	Latent trait	.965	.103	.280	.046	.646
	Discrimination	.664	-.011	.151	.034	.512
	Difficulty	.995	.161	.172	.034	.528
	Guessing	-.030	.004	.030	.004	.269
Cholesky	Latent trait	.952	-.039	.311	.127	.735
	Discrimination	.691	-.106	.210	.058	.598
	Difficulty	.987	-.066	.189	.128	.657
	Guessing	.106	-.001	.029	.004	.274

Table 5: Models comparison: Simulation study

Model	DIC	EAIC	EBIC
Copula	<b>37326.58</b>	<b>38049.87</b>	<b>44191.53</b>
Cholesky	37998.86	39629.03	53471.32

Table 6: Structure of the tests: real data analysis

	Test 1	Test 2	Test 3
Test 1	27	6	5
Test 2	6	34	9
Test 3	5	9	30

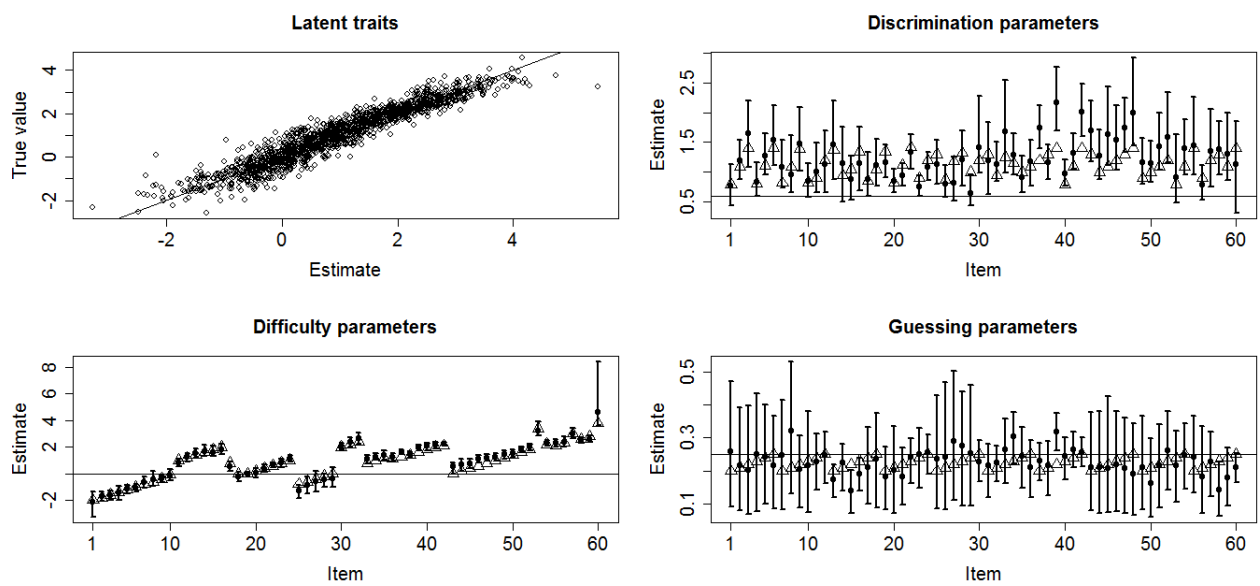


Figure 3: Estimates of latent traits and item parameters with 95% credibility intervals under the Cholesky model. Legend: Circles denotes estimates, triangles denotes true values.

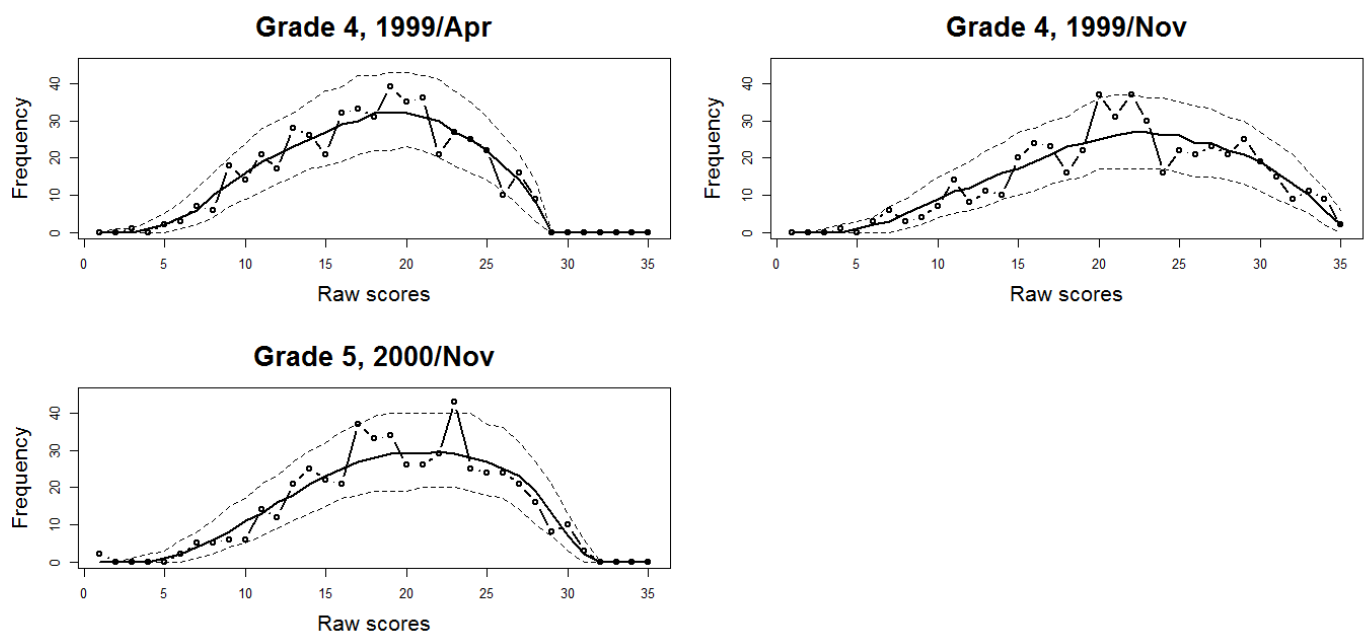


Figure 4: Observed and predicted scores distributions with 95% credibility intervals for the copula model

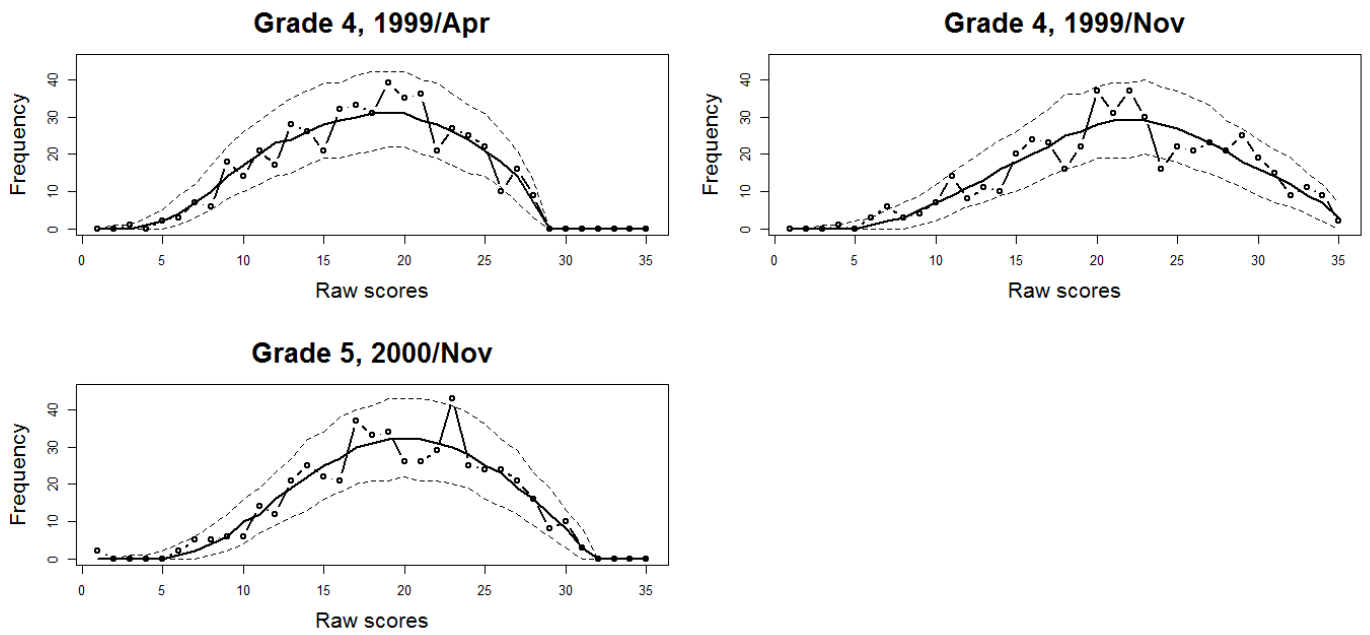


Figure 5: Observed and predicted scores distributions with 95% credibility intervals for Cholesky model

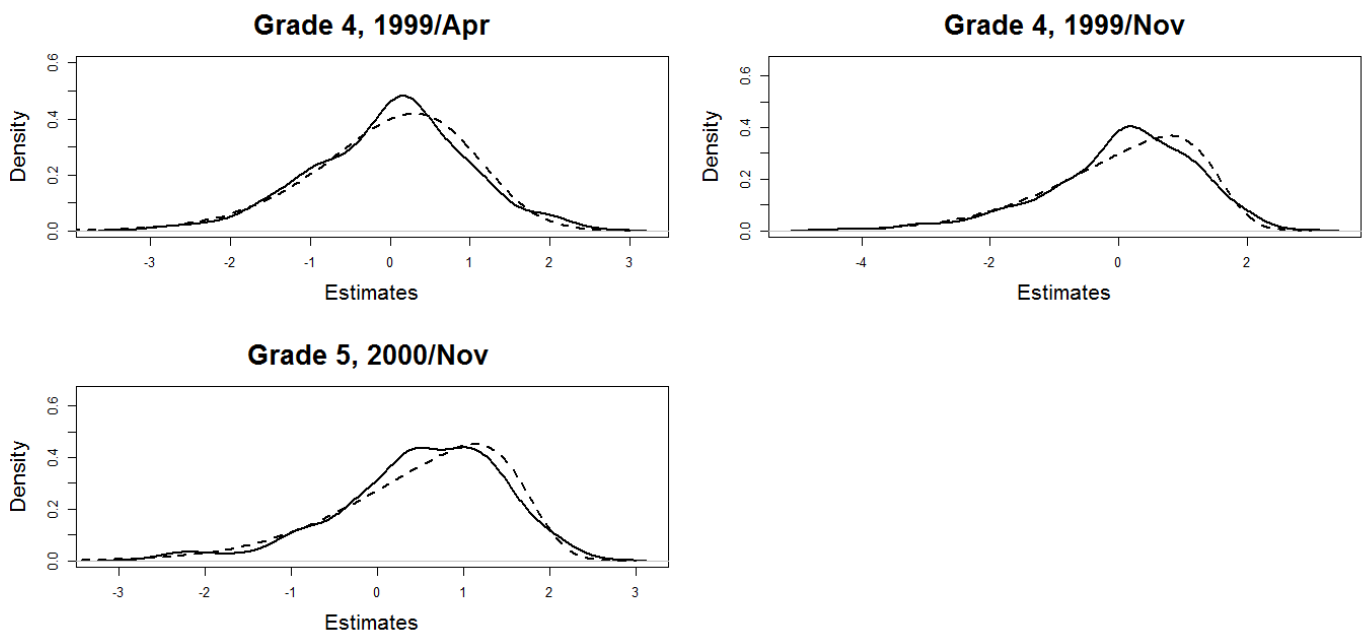


Figure 6: Estimated latent traits distributions with theoretical curves. Legend: Smoothed histogram (solid line), copula model (dashed line)

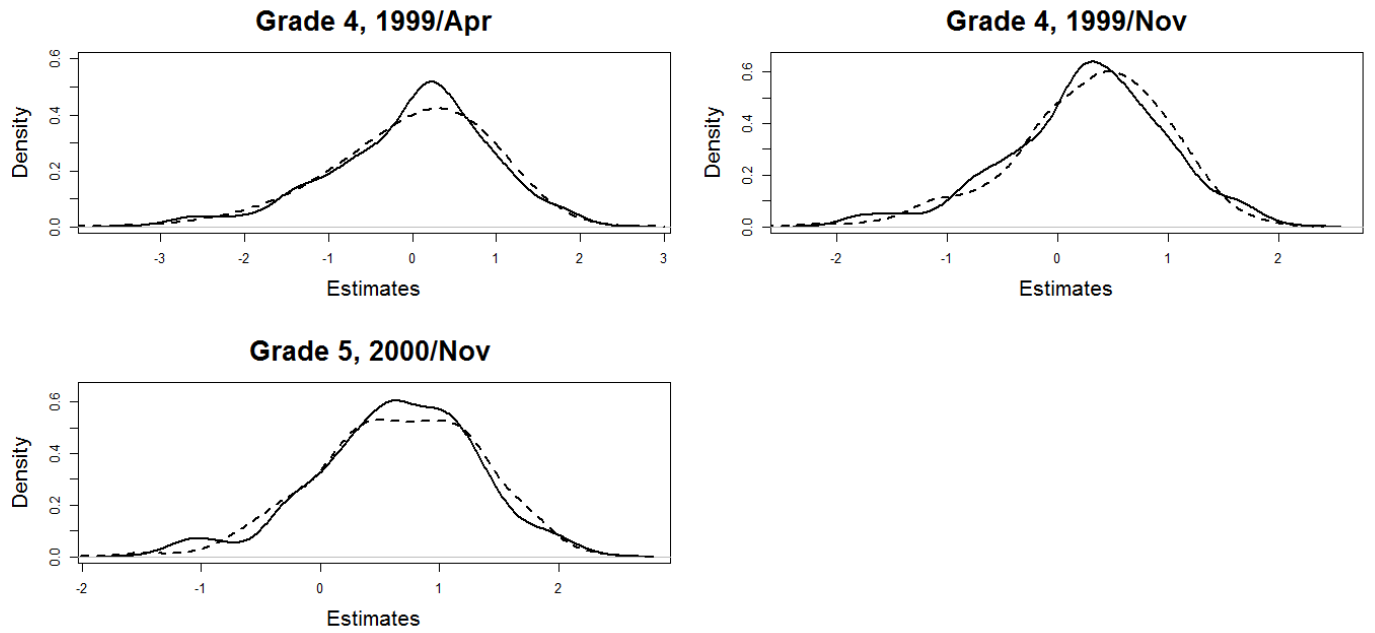


Figure 7: Estimated latent traits distributions with theoretical curves. Legend: Smoothed histogram (solid line), Cholesky model (dashed line)

Table 7: Estimates of the population parameters considering both copula and Cholesky models

	Copula			Cholesky		
	Mean	SD	CI (95%)	Mean	SD	CI (95%)
$\mu_{\theta_1}$	.000	—	—	.000	—	—
$\mu_{\theta_2}$	.053	.104	[-.124, .252]	.259	.059	[.133, .375]
$\mu_{\theta_3}$	.526	.091	[.332, .681]	.631	.057	[.518, .740]
$\sigma_{\theta_1}^2$	1.000	—	—	1.000	—	—
$\sigma_{\theta_2}^2$	1.486	.344	[.978, 2.277]	.548	.117	[.346, .822]
$\sigma_{\theta_3}^2$	.976	.155	[.702, 1.286]	.503	.144	[.284, .852]
$\rho_{\theta_1}$	.810	.018	[.778, .853]	.847	.020	[.807, .882]
$\rho_{\theta_2}$	.856	.029	[.797, .912]	.851	.024	[.800, .895]
$\gamma_{\theta_1}$	-.500	.242	[-.858, .032]	-.526	.127	[-.743, -.229]
$\gamma_{\theta_2}$	-.823	.128	[-.988, -.530]	-.366	.090	[-.519, -.158]
$\gamma_{\theta_3}$	-.827	.113	[-.983, -.558]	-.319	.063	[-.433, -.188]



Figures 8 and 9 present the Bayesian  $p$ -value for the item parameters considering the two models. Items with  $p$ -value below .05 or above .90 were considered to be not well fitted. The results are similar and indicate that the two model fitted the items properly, with few exceptions. Figures 10 to 12 present the estimates of the item parameters considering the copula and Cholesky models. The circles and triangles, denote the estimates via copula and Cholesky model, respectively, and vertical bar represent the 95% credibility intervals. We can see that, the discrimination parameters estimates tends to be higher under the Cholesky model. The difficulty and guessing parameters estimates are similar for the most of the items.

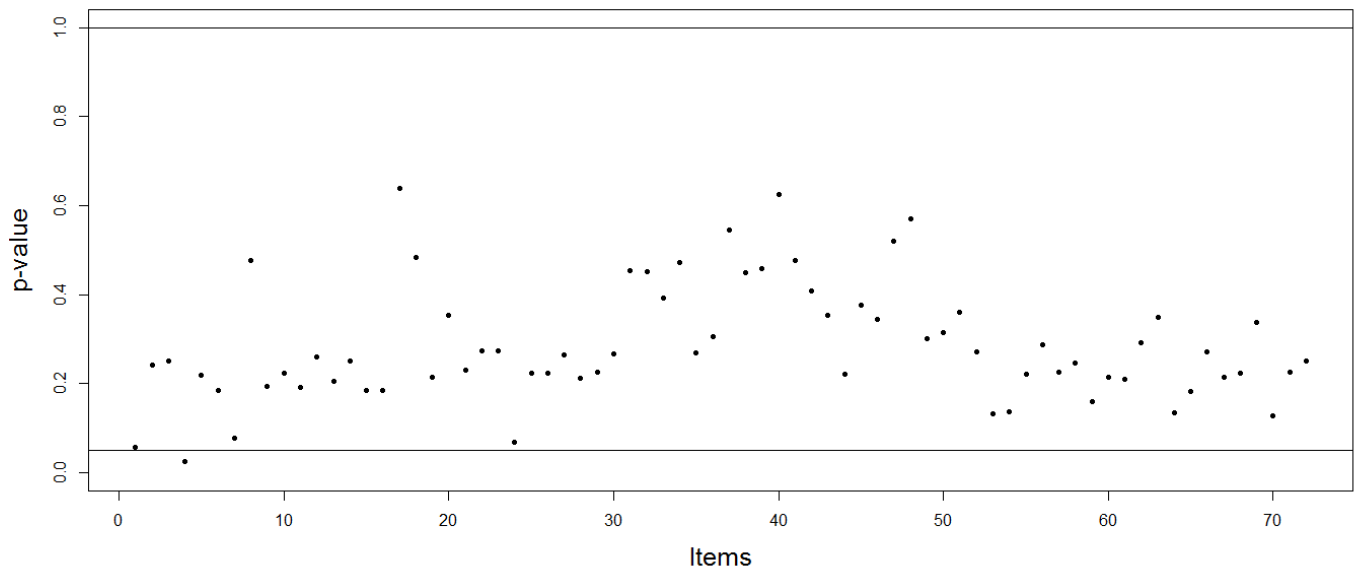


Figure 8: Bayesian  $p$ -values for item parameters considering the copula model

Lastly, we compare the models in terms of the global adjustment by using the comparison statistics. Most of the statistics selected the copula model.

Table 8: Models comparison: Real data analysis

	DIC	EAIC	EBIC
Copula	54301.44	<b>54633.80</b>	<b>65283.64</b>
Cholesky	<b>54267.23</b>	54643.03	68541.70

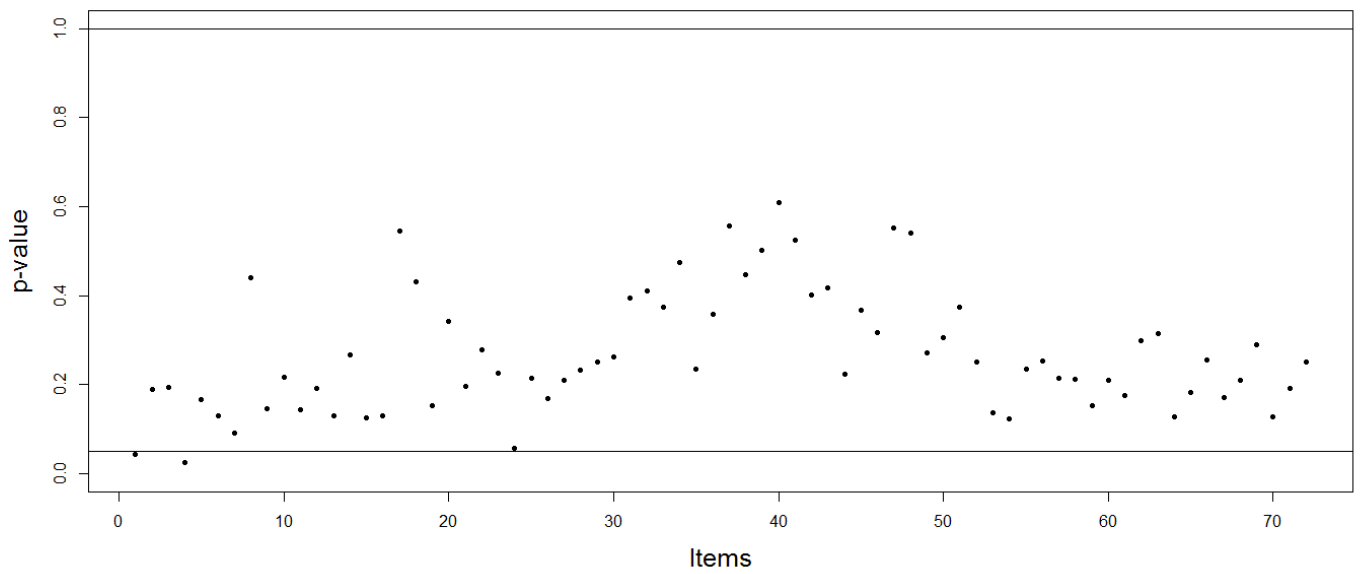


Figure 9: Bayesian  $p$ -values for item parameters considering the Cholesky model

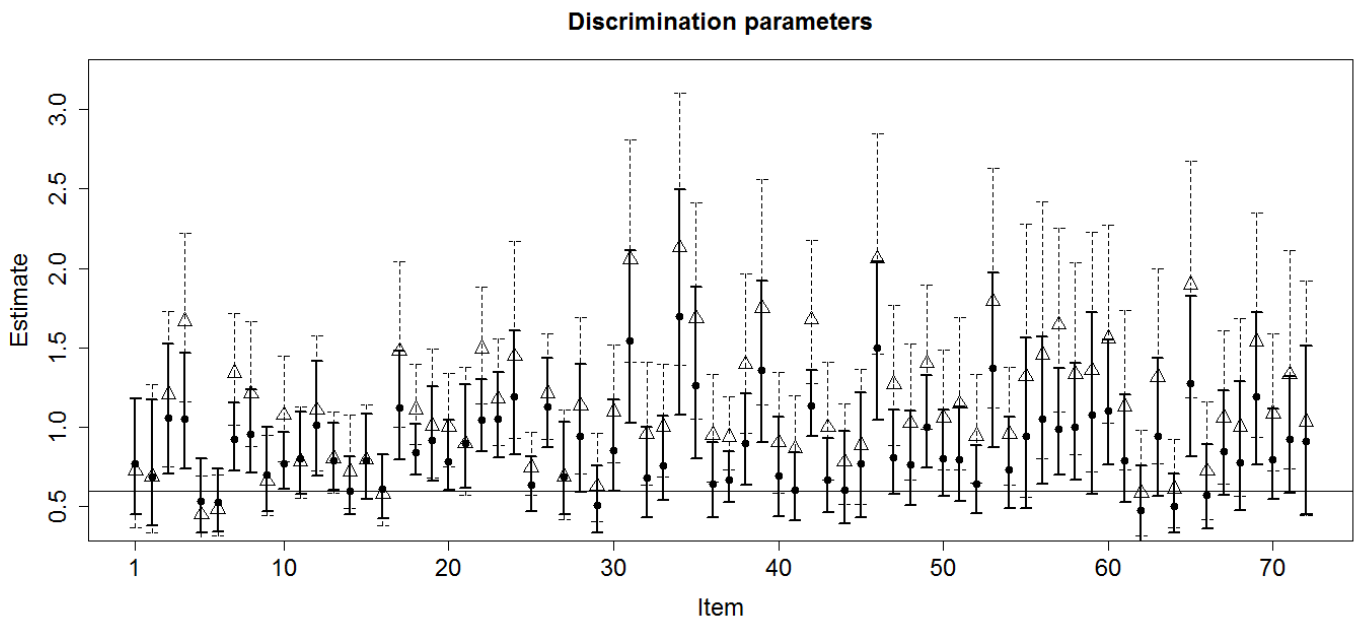


Figure 10: Discrimination parameters estimates.

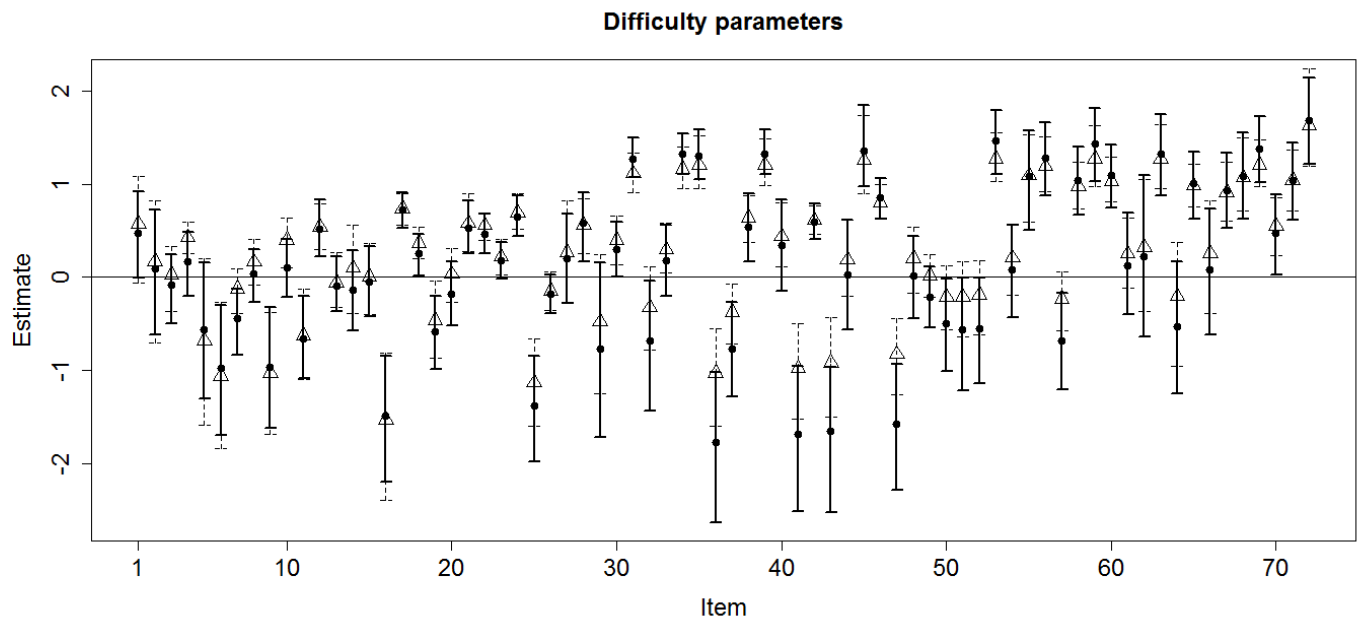


Figure 11: Difficulty parameters estimates.

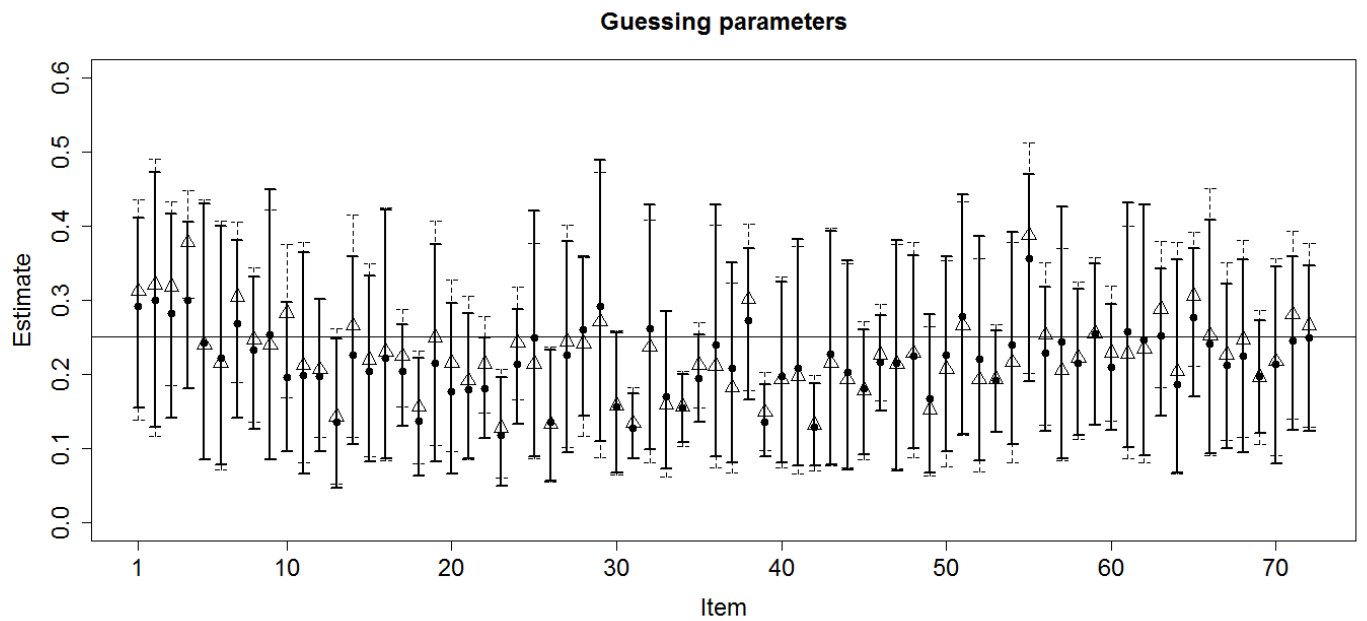


Figure 12: Guessing parameters estimates.

## 6 Concluding Remarks

We presented a copula based IRT longitudinal model with skewed latent trait distributions. Such model shows to a very promising alternative in analysis of longitudinal IRT data. It allows to consider a wide range of specific correlation structures through the using of copula functions. This methodology also provides an absolute control of the marginal latent trait distributions. However, the Cholesky decomposition based modeling seems to be more flexible to incorporate dropouts and/or inclusion of subjects in the modeling. An MCMC algorithm to estimate the parameters of the model was developed. It is a hybrid algorithm that combines Metropolis-Hastings with SVE steps within a Gibbs sampling framework. The SVE algorithm with oversampling was built for both correlation and latent trait parameters simulation and it showed to be a good alternative to Metropolis-Hastings algorithm in terms of convergence and quality of proposal values. A simulation study was conducted to compare the skew copula model and the skew Cholesky decomposition based model in terms of parameter recovery. We could see that the copula model tends to be more efficient, in terms of parameter recovery, especially to estimate the latent traits, discrimination and skewness parameters, when data is generated from the copula model. Furthermore, a real data set concerning to a Brazilian school development study was analyzed. Some model fit assessment tools were considered, indicating that the copula model was well fitted. The copula model also outperforms the Cholesky model in this analysis according to some model comparison criteria. In conclusion, our approach revealed to be a promising alternative for analyzing longitudinal IRT data. In future research we intend to explore some extensions of this model in order to consider other copula functions and other marginal distributions. Alternative algorithms could also be considered in order to reduce computational effort.

## 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. [12](#)
- 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. [16](#)
- 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. [19](#)
- Azzalani, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 12:171–178. [4](#)
- Bedford, T. and Cooke, R. (2002). Vines: a new graphical model for dependent random variables. *Annals of Statistics*, 30:1031–1068. [2](#)
- 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. [7](#)
- Braeken, J., Tuerlinckx, F., and Boeck, P. (2007). Copula functions for residual dependency. *Psychometrika*, 72(3):393–411. [2](#)
- Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*,. Wiley, Chichester, England, 1 edition. [2](#)
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151. [2](#)
- Danaher, P. and Smith (2011). Modeling multivariate distributions using copulas: applications in marketing. *Marketing Science*, 30(1):4–21. [2](#)
- Doebler, A. and Doebler, P. (2012). Building partially compensatory mirt models with copulae. working paper. [2](#)

- Frahm, G., Junker, M., and Szimayer, A. (2003). Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286. 2, 5
- Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25. 2
- Jiafeng, S., Edward, W. F., and R., M. A. (2008). Heavy-tailed longitudinal data modeling using copulas. *Insurance: Mathematics and Economics*, 42(2):817 – 830. 2
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, 1 edition. 5
- Li, D. X. (1999). On default correlation: A copula function approach. *The RiskMetrics Group*. 2
- Montgomery, D. C. (2004). *Design and Analysis of Experiments*. Chapman & Hall, London, sixth edition. 8
- Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer, 2 edition. 2, 3, 5
- Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493. 2
- Pewsey, A. (2000). Problems of inference for azzalini’s skewnormal distribution. *Journal of Applied Statistics*, 27(7):859–870. 4
- Reckase, M. (2009). *Multidimensional Item Response Theory*. Springer New York, 1 edition. 2
- Santos, J. (2016). Contributions to the study of longitudinal item response theory data. PhD Thesis. 14
- Santos, J., Azevedo, C., and Fox, J. (2017a). A general cholesky decomposition based modeling of longitudinal irt data. Technical report. 1, 13, 14, 19
- Santos, J., Azevedo, C., and Fox, J. (2017b). A general cholesky decomposition based modeling of longitudinal irt data: Handling skewed latent traits distributions. Technical report. 1, 2, 4, 15, 19

- 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. 16
- Skalar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de l’Institut de Statistique de L’Universite de Paris (In French)*, 8:229–231. 3
- Smith, M., Min, A., Almeida, C., and Czado, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105(492):1467–1479. 2, 5
- 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. 18
- 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
- Xue-Kun Song, P. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320. 5