



UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística
e Computação Científica

JOSÉ ROBERTO SILVA DOS SANTOS

CONTRIBUIÇÕES AO ESTUDO DE DADOS LONGITUDINAIS NA
TEORIA DE RESPOSTA AO ÍTEM

CONTRIBUTIONS TO THE STUDY OF LONGITUDINAL ITEM
RESPONSE THEORY DATA

CAMPINAS
2016



JOSÉ ROBERTO SILVA DOS SANTOS

CONTRIBUTIONS TO THE STUDY OF LONGITUDINAL ITEM
RESPONSE THEORY DATA

CONTRIBUIÇÕES AO ESTUDO DE DADOS LONGITUDINAIS NA
TEORIA DE RESPOSTA AO ÍTEM

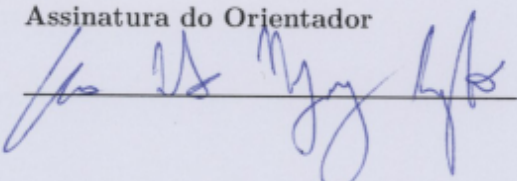
Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Estatística.

Thesis presented to the Institute of Mathematics, Statistics and Scientific Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Statistics.

Orientador: Caio Lucidius Naberezny Azevedo

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO JOSÉ ROBERTO SILVA DOS SANTOS, E ORIENTADA PELO PROF. DR. CAIO LUCIDIUS NABEREZNY AZEVEDO.

Assinatura do Orientador

A handwritten signature in blue ink is written over a horizontal line. The signature is cursive and appears to read "Caio Naberezny Azevedo".

CAMPINAS
2016

Agência(s) de fomento e nº(s) de processo(s): CAPES; CNPq, 162562/2014-4 e 142486/2015-9

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Maria Fabiana Bezerra Muller - CRB 8/6162

Santos, José Roberto Silva dos, 1984-
Sa59c Contributions to the study of longitudinal item response theory data / José Roberto Silva dos Santos. – Campinas, SP : [s.n.], 2016.

Orientador: Caio Lucidius Naberezny Azevedo.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Teoria da resposta ao item. 2. Estudos longitudinais. 3. Inferência bayesiana. 4. Métodos MCMC. I. Azevedo, Caio Lucidius Naberezny, 1979-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Contribuições ao estudo de dados longitudinais na teoria de resposta ao item

Palavras-chave em inglês:

Item response theory

Longitudinal studies

Bayesian inference

MCMC methods

Área de concentração: Estatística

Titulação: Doutor em Estatística

Banca examinadora:

Caio Lucidius Naberezny Azevedo [Orientador]

Filidor Edilfonso Vilca Labra

Dani Gamerman

Hedibert Freitas Lopes

Jorge Luis Bazán

Data de defesa: 25-10-2016

Programa de Pós-Graduação: Estatística

Tese de Doutorado defendida em 25 de outubro de 2016 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). CAIO LUCIDIUS NABEREZNY AZEVEDO

Prof(a). Dr(a). FILIDOR EDILFONSO VILCA LABRA

Prof(a). Dr(a). DANI GAMERMAN

Prof(a). Dr(a). HEDIBERT FREITAS LOPES

Prof(a). Dr(a). JORGE LUIS BAZÁN

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do aluno.

*Aos meus familiares Helena, Ribamar e Wesckley.
E à minha namorada Sandra.*

Para minha mãe Helena

*"Os teus olhos são estrelas que aveludam
meu coração
São as estrelas mais belas que eu vi em
noites de verão.
Eu fico encantado por esse brilho incan-
descendente
Dão luz à minha vida, iluminam a minha
mente.*

*Trazem-me muita alegria para feliz poder
viver
Esse olhar é sinfonia para nunca mais
esquecer.
Quando acordas pela manhã, teus olhos
brilham tanto
Que os meus ficam inundados por esse
teu encanto.*

*Durante o dia esse brilho, ilumina os
meus passos.
São a luz da minha vida, são a força dos
meus braços.
Se o meu corpo tem calor, a ti ele te deve
também
Olhos que me acompanham, são os teus
olhos minha mãe.*

*Olhos de encantar.
Alberto da Fonseca.*

Para Sandra

*De almas sinceras a união sincera
Nada há que impeça: amor não é amor
Se quando encontra obstáculos se altera,
Ou se vacila ao mínimo temor.
Amor é um marco eterno, dominante,
Que encara a tempestade com bravura;
É astro que norteia a vela errante,
Cujo valor se ignora, lá na altura.
Amor não teme o tempo, muito embora
Seu alfange não poupe a mocidade;
Amor não se transforma de hora em hora,
Antes se afirma para a eternidade.
Se isso é falso, e que é falso alguém
provou,
Eu não sou poeta, e ninguém nunca amou.*

William Shakespeare.

Agradecimentos

Gostaria de nas próximas linhas lembrar daqueles que fizeram parte desta trajetória e que fazem parte da minha vida. Dificilmente, conseguiria citar tão brevemente todos aqueles que, de alguma forma, contribuíram para conclusão desta etapa. Porém, as pessoas aqui lembradas representam todos a quem devo minha eterna gratidão. Desta forma, agradeço com respeito e sinceridade à:

Deus, pela dádiva da vida, pelo seu inimaginável amor e por ter me concedido saúde e paz para o desenvolvimento deste trabalho.

Aos meus pais, Helena e Ribamar, pelo amor incondicional, pelo apoio em qualquer situação, pelos profundos ensinamentos e orientações, por conduzirem as nossas vidas com sabedoria e dedicação. Em especial à minha mãe pela alegria e conforto de suas palavras, pela total doação e dedicação para com nossa educação, pelas atitudes e orações, pelo seu incansável zelo e vigília com a mão que nunca parou de me ninar e pela segurança que sempre me passou e que me fez chegar até aqui.

Ao meu irmão Wesckley pelo apoio a qualquer momento, pelas brincadeiras fraternas e pelas calorosas recepções a cada reencontro.

À toda minha família, pelo carinho, amor e por compreenderem os momentos de ausência.

À minha namorada Sandra, pelo carinho, ternura, atenção e desprendimento. Por ter compartilhado comigo diversos momentos (de alegrias e de incertezas), através dos quais, aprendemos um com o outro e fortalecemos a nossa união. Pela simplicidade cativante, pelos lindos sonhos compartilhados e por tudo mais que já não poderia expressar, como diria Dante Alighieri: "Muito pouco ama, quem com palavras pode expressar quanto muito ama."

À todos os meus amigos de Fortaleza-CE, especialmente à Tiago, William ("Berada"), Junior ("Chumbim"), pela longa amizade e respeito, por todos os momentos felizes (desde a infância) e pelas calorosas recepções a cada reencontro.

À todos os amigos, professores e funcionários do DEMA-UFC. Em especial aos amigos Dival, William e Mota, pela ajuda, apoio e torcida, pelas brincadeiras e momentos de descontração. Aos professores Maurício, Juvêncio, Ailton, Sílvia e Ana Maria, pelos conhecimentos repassados, atenção, incentivo e por acreditarem em minha dedicação. Em especial, agradeço ao grande professor Maurício por sua amizade, ajuda e conselhos que me ajudaram bastante a escolher meu caminho após a graduação. Agradeço também, as funcionárias Márgeri, Luísa e

Mariluse que sempre nos atenderam com muita atenção e simpatia.

À todos os amigos do IMECC-UNICAMP, pela ótima convivência, ajuda, churrascos e diversos outros momentos agradáveis. Em especial agradeço aos amigos Carlos, Larissa, Marcos, Diana, José Alejandro, Julian, Juan Leonardo, Ali e Nicolás.

I would like to thank the OMD department of the University of Twente to host me in my internship abroad. Specially I would like to say thank to my supervisor Prof. Jean-Paul Fox, for the many interesting discussions we had and for making me feel welcome. I thank secretary Mrs. Bosch for all support with paperwork. I also thank my colleagues Nabeel and Muirne, for the city tours (cycling our essential *fietzen*), dinners and for all support.

Ao Sr. Otávio, pelo excelente convivência em vossa pensão, pela solicitude e hospitalidade.

Ao meu orientador Caio, pela amizade, apoio, respeito e confiança.

Ao Professor Héilton Tavares pela participação em nossa banca de qualificação, pela solicitude e por todos os esclarecimentos fornecidos a respeito do conjunto de dados PDE analisado nesta tese.

Aos membros da banca examinadora Professores Hedibert Lopes, Dani Gamerman, Jorge Bazán e Filidor Labra, pelas sugestões, críticas e contribuições que deram ao presente trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte financeiro.

Resumo

Na presente tese desenvolvemos classes de modelos longitudinais da Teoria de Resposta o Item (TRI) considerando duas abordagens. A primeira é baseada na decomposição de Cholesky de matrizes de covariância de interesse, relacionadas aos traços latentes. Essa metodologia permite representar um amplo conjunto de estruturas de dependência de maneira relativamente simples, facilita a escolha de distribuições a priori para os parâmetros relacionados à estrutura de dependência, facilita a implementação de algoritmos de estimação (particularmente sob o enfoque Bayesiano), permite considerar diferentes distribuições (multivariadas) para os traços latentes de modo simples, torna bastante fácil a incorporação de estruturas de regressão para os traços latentes, entre outras vantagens. Desenvolvemos, adicionalmente, uma classe de modelos com estruturas de curvas de crescimento para os traços latentes. Na segunda abordagem utilizamos cópulas Gaussianas para representar a estrutura de dependência dos traços latentes. Diferentemente da abordagem anterior, essa metodologia permite o total controle das respectivas distribuições marginais mas, igualmente, permite considerar um grande número de estruturas de dependência. Utilizamos modelos dicotômicos de resposta ao item e exploramos a utilização da distribuição normal e normal assimétrica para os traços latentes. Consideramos indivíduos acompanhados ao longo de várias condições de avaliação, submetidos a instrumentos de medida em cada uma delas, os quais possuem alguma estrutura de itens comuns. Exploramos os casos de um único e de vários grupos como também dados balanceados e desbalanceados, no sentido de considerarmos inclusão e exclusão de indivíduos ao longo do tempo. Algoritmos de estimação, ferramentas para verificação da qualidade de ajuste e comparação de modelos foram desenvolvidos sob o paradigma bayesiano, através de algoritmos MCMC híbridos, nos quais os algoritmos SVE (Single Variable Exchange) e Metropolis-Hastings são considerados quando as distribuições condicionais completas não são conhecidas. Estudos de simulação foram conduzidos, os quais indicaram que os parâmetros foram bem recuperados. Além disso, dois conjuntos de dados longitudinais psicométricos foram analisados para ilustrar as metodologias desenvolvidas. O primeiro é parte de um estudo de avaliação educacional em larga escala promovido pelo governo federal brasileiro. O segundo foi extraído do Amsterdam Growth and Health Longitudinal Study (AGHLS) que monitora a saúde e o estilo de vida de adolescentes holandeses.

Palavras-chave: Teoria da Resposta ao Item, estudos longitudinais, Inferência bayesiana, métodos MCMC, decomposição de Cholesky, cópulas (Estatística matemática).

Abstract

In this thesis we developed families of longitudinal Item Response Theory (IRT) models considering two approaches. The first one is based on the Cholesky decomposition of the covariance matrices of interest, related to the latent traits. This modeling can accommodate several dependence structures in a easy way, it facilitates the choice of prior distributions for the parameters of the dependence matrix, it facilitates the implementation of estimation algorithms (particularly under the Bayesian paradigm), it allows to consider different (multivariate) distributions for the latent traits, it makes easier the inclusion of regression and multilevel structures for the latent traits, among other advantages. Additionally, we developed growth curve models for the latent traits. The second one uses a Gaussian copula function to describes the latent trait structure. Differently from the first one, the copula approach allows the entire control of the respective marginal latent trait distributions, but as the first one, it accommodates several dependence structures. We focus on dichotomous responses and explore the use of the normal and skew-normal distributions for the latent traits. We consider subjects followed over several evaluation conditions (time-points) submitted to measurement instruments which have some structure of common items. Such subjects can belong to a single or multiple independent groups and also we considered both balanced and unbalanced data, in the sense that inclusion or dropouts of subjects are allowed. 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, two longitudinal psychometrical data sets were analyzed to illustrate our methodologies. The first one is a large-scale longitudinal educational study conducted by the Brazilian federal government. The second was extracted from the Amsterdam Growth and Health Longitudinal Study (AGHLS), which monitors the health and life-style of Dutch teenagers.

Keywords: Item Response Theory, Longitudinal studies, Bayesian inference, MCMC methods, Cholesky decomposition, Copula modeling.

List of Figures

2.1	Posterior means and 95% credibility intervals for item parameters for ARH(1) model under balanced data. Legend: Circles (estimates), triangles (true values) and Bars (credibility intervals)	46
2.2	Posterior means and 95% credibility intervals for item parameters for ARH(1) model under unbalanced data. Legend: Circles (estimates), triangles (true values) and Bars (credibility intervals)	46
2.3	Correlation profiles. Legend: unstructured matrix (\bullet), structured matrices ($-\Delta-$)	49
2.4	Observed score distribution and expected score distribution, and 95% central credibility intervals. Legend: observed scores (dotted line), expected frequency (solid line) and central credibility interval (dashed line)	50
2.5	Estimated latent traits distributions. Legend: Smoothed histograms (dashed line), Theoretical curve (Solide line)	51
2.6	Bayesian p -values for item parameters	52
2.7	Posterior means and 95% central credibility intervals for discrimination parameters	53
2.8	Posterior means and 95% central credibility intervals for difficulty parameters . .	53
2.9	Posterior means and 95% central credibility intervals for guessing parameters . .	54
2.10	Observed score distribution and expected score distribution, and 95% central credibility intervals. Legend: observed scores (dotted line), expected frequency (solid line) and central credibility interval (dashed line). Unbalanced case. . . .	54
2.11	Estimated latent traits distributions (Unbalanced case). Legend: Smoothed histograms (dashed line), Theoretical curve (Solide line)	55
2.12	Bayesian p -values for item parameters (Unbalanced case).	56
2.13	Posterior means and 95% central credibility intervals discrimination parameters (Unbalanced case).	56
2.14	Posterior means and 95% central credibility intervals difficulty parameters (Unbalanced case).	57
2.15	Posterior means and 95% central credibility intervals guessing parameters (Unbalanced case).	57
3.1	Prior distributions of the population parameters	77
3.2	Estimates of latent traits and item parameters. Legend: circles denotes the estimates, triangles denotes the true values and the vertical bars denote 95% credibility intervals	79
3.3	Correlation profiles. Legend: unstructured matrix (\bullet), structured matrices ($-\Delta-$).	83
3.4	Observed and predicted scores distributions with 95% credibility intervals	83
3.5	Estimated latent traits distributions according to the symmetric model. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)	85

3.6	Estimated latent traits distributions according to the skew model. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)	85
3.7	Bayesian p -values for the items	86
3.8	Posterior means and 95% central credibility intervals for discrimination parameters.	86
3.9	Posterior means and 95% central credibility intervals for difficulty parameters.	87
3.10	Posterior means and 95% central credibility intervals for guessing parameters.	87
4.1	Underlying growth curves for the simulation study. Legend: Group 1 (solid line), Group 2 (dashed line)	103
4.2	Estimates of latent traits and item parameters. Legend: circles denotes estimates, triangles denotes true values and vertical bars denote 95% credibility intervals	106
4.3	Estimates of the population means considering the Multiple Group Model. Legend: Female ($- \circ -$), Male ($- \Delta -$)	109
4.4	Estimated growth curves. Legend: Female (solid line), Male (dashed line)	110
4.5	Estimated means with 95% central credibility intervals. Legend: Female ($- \circ -$), Male ($- \Delta -$).	110
4.6	Observed and predicted scores distributions with 95% credibility intervals for female students	111
4.7	Observed and predicted scores distributions with 95% credibility intervals for male students	112
4.8	Estimated latent trait distributions for female students. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)	113
4.9	Estimated latent trait distributions for male students. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)	113
4.10	Estimated difference between the group means along the years. Legend: Estimate (circles), 95% credibility intervals (vertical bars)	114
4.11	Bayesian p -values for item parameters	115
4.12	Item parameters estimates. Legend: estimates (circles), 95% credibility intervals (vertical bars)	115
5.1	Estimated latent traits distributions. Legend: True distributions (solid line), Copula model (dashed line) and Cholesky model (dotted line)	130
5.2	Estimates of latent traits and item parameters with 95% credibility intervals under the copula model. Legend: Circles denotes estimates, triangles denotes true values.	131
5.3	Estimates of latent traits and item parameters with 95% credibility intervals under the Cholesky model. Legend: Circles denotes estimates, triangles denotes true values.	134
5.4	Observed and predicted scores distributions with 95% credibility intervals for the copula model	134
5.5	Observed and predicted scores distributions with 95% credibility intervals for Cholesky model	135
5.6	Estimated latent traits distributions with theoretical curves. Legend: Smoothed histogram (solid line), copula model (dashed line)	135
5.7	Estimated latent traits distributions with theoretical curves. Legend: Smoothed histogram (solid line), Cholesky model (dashed line)	136

5.8	Bayesian p -values for item parameters considering the copula model	137
5.9	Bayesian p -values for item parameters considering the Cholesky model	137
5.10	Discrimination parameters estimates.	138
5.11	Difficulty parameters estimates.	138
5.12	Guessing parameters estimates.	139

List of Tables

2.1	Structured covariance matrices used in this work.	30
2.2	Effective sample size and effective sample size per hour considering two latent traits samplers	41
2.3	Correlation parameters for each dependence structure	42
2.4	Hyperparameters for the prior distribution	42
2.5	Results for the population means parameters	43
2.6	Results for the population variance parameters	43
2.7	Results for the population correlation parameters.	44
2.8	Results for the estimated latent traits and item parameters.	44
2.9	Results for the population parameters for the unbalanced data study	45
2.10	Results for the latent traits and item parameters for the unbalanced data study	45
2.11	Structure of tests: real data analyze	48
2.12	Models comparison: Real data analysis	49
2.13	Estimates of the population parameters according to the AD model	52
2.14	Estimates of the population parameters according to the AD model (Unbalanced case)	55
3.1	Structured covariance matrices used in this work. The σ -parameters are related to variances, while ρ -parameters are related to correlations.	65
3.2	Hyperparameters for the prior distributions	77
3.3	Results for the population parameters.	78
3.4	Results for the estimated latent traits and item parameters.	79
3.5	Structure of tests: real data analyze	81
3.6	Statistics for model comparison	82
3.7	Estimates of the population parameters according to the AD model considering both symmetric and skew models	84
4.1	Structured covariance matrices used in this work. The σ -parameters are related to variances, while ρ -parameters are used for correlations.	93
4.2	Growth curves for population means	94
4.3	Restricted growth curves for population means	94
4.4	Item parameters for the recovery parameters study	104
4.5	The priors distributions hyperparameters	104
4.6	Population parameters estimates of the Group 1	105
4.7	Population parameters estimates of the Group 2	105
4.8	Results for the estimated latent traits and item parameters.	106
4.9	Number of subjects according to gender	108
4.10	Estimated correlation structure for female and male group	109

4.11	Statistics for comparison of the correlation structures	109
4.12	Statistics for comparison of the growth curves	109
4.13	Estimates of the population parameters for the female students	112
4.14	Estimates of the population parameters for the male students	114
5.1	Structured correlation matrices used in this work.	121
5.2	The priors distributions hyperparameters	129
5.3	Results for the estimated population parameters	133
5.4	Results for the estimated latent traits and item parameters.	133
5.5	Models comparison: Simulation study	133
5.6	Structure of the tests: real data analysis	133
5.7	Estimates of the population parameters considering both copula and Cholesky models	136
5.8	Models comparison: Real data analysis	136
B.1	Item parameters of the simulation studies of the Chapters 2 and 3	152
B.2	Item parameters of the simulation study of the Chapter 5	153

List of Algorithms

2.1	Full Gibbs sampling algorithm	35
2.2	FFBS algorithm	36
2.3	The Single-Variable Exchange algorithm	37
2.4	The Single-Variable Exchange algorithm with Oversampling	38
2.5	The SVE algorithm with oversampling to sample a correlation parameter considering ARH(1) matrix	39
2.6	Gibbs sampling with FFBS sampler for unstructured matrix	40
2.7	Gibbs sampling with FFBS and SVE samplers for structured matrices	40
3.1	FFBS algorithm	73
3.2	The SVE algorithm with oversampling to sample a correlation parameter considering ARH(1) matrix	74
3.3	Gibbs sampling with FFBS for unstructured matrix	75
3.4	Gibbs sampling with FFBS for structured matrices	75
4.1	Growth curve parameters sampler	98
4.2	Variance parameters sampler	98
4.3	Conditional skewness parameters sampler	98
4.4	FFBS algorithm	100
4.5	The SVE algorithm with oversampling to sample a correlation parameter considering ARH(1) matrix	101
4.6	General algorithm for unstructured matrix	102
4.7	General algorithm for structured matrices	102
5.1	Population mean sampler	124
5.2	Population variance sampler	124
5.3	Population skewness coefficient sampler	125
5.4	The SVE algorithm with oversampling to sample a correlation parameter considering AR(1) matrix	127
5.5	The SVE algorithm with oversampling to sample latent traits	128
5.6	Full Sampler for the skew copula IRT longitudinal model	128
A.1	The SVE algorithm with oversampling to sample a correlation parameter considering HT matrix	148

Contents

1	Introduction	20
1.1	Overview	20
1.2	Motivation and literature review	20
1.3	Objectives and outline	24
1.3.1	Objectives	24
1.3.2	Outline	24
2	Bayesian general Cholesky decomposition based modeling of longitudinal IRT data.	25
2.1	Introduction	26
2.2	Modeling	27
2.2.1	Antedependence models	28
2.3	Bayesian Estimation and MCMC Algorithms	31
2.3.1	Prior and posterior distributions	32
2.3.2	An alternative sampler for the latent traits	35
2.3.3	Oversampling in the Single-Variable Exchange algorithm	37
2.4	Simulation Studies	40
2.4.1	Parameter Recovery	41
2.5	Real Data Analysis and Model fit Assessment	47
2.5.1	Model fit assessment tools	47
2.5.2	Model Comparison	47
2.5.3	Brazilian school development study	48
2.6	Concluding Remarks	58
3	Bayesian general Cholesky decomposition based modeling of longitudinal IRT data with skewed latent distributions.	59
3.1	Introduction	60
3.2	The Skew-Normal Distribution Under the Centered Parametrization	61
3.3	Modeling	62
3.3.1	Antedependence models	63
3.3.2	Model identification	67
3.4	Bayesian Estimation and MCMC Algorithms	67
3.4.1	Prior specification and posterior distributions	69
3.4.2	The latent traits sampling	71
3.4.3	An alternative correlation parameters sampler	73
3.5	Parameter recovery study	76

3.6	Real Data Analysis and Model fit Assessment	80
3.6.1	Model fit assessment tools	80
3.6.2	Model Comparison	80
3.6.3	The Brazilian school development study	81
3.7	Concluding Remarks	88
4	Bayesian general Cholesky decomposition based modeling of longitudinal multiple-group IRT data with skewed latent distributions and growth curves.	89
4.1	Introduction	90
4.2	Modeling	90
4.2.1	Latent traits modeling	91
4.2.2	Growth curves for population means	93
4.3	Bayesian Estimation and MCMC Algorithms	94
4.3.1	Prior and posterior distributions	95
4.4	Parameter recovery study	103
4.5	Real Data Analysis and Model fit Assessment	107
4.5.1	Model fit assessment tools	107
4.5.2	Model Comparison	107
4.5.3	Amsterdam Growth and Health Longitudinal Study	108
4.6	Concluding Remarks	116
5	A Copula Based Modeling for Longitudinal IRT Data with skewed latent distributions.	117
5.1	Introduction	118
5.2	Modeling	118
5.3	Bayesian Estimation and MCMC Algorithms	121
5.3.1	Prior and posterior distributions	122
5.3.2	Correlation parameters sampler	126
5.3.3	Latent trait parameters estimation	127
5.4	Simulation Study	127
5.5	The Brazilian school development study	130
5.6	Concluding Remarks	140
6	Final Conclusions	141
	Bibliography	143
A	Further results on SVE algorithms	148
A.1	SVE algorithm for Heteroscedastic Toeplitz matrix	148
A.2	Some useful R functions	149
A.2.1	Cholesky decomposition	149
A.2.2	Functions to build structured matrices	149
B	Item parameters of the simulation studies	151

Chapter 1

Introduction

1.1 Overview

In many research fields, as psychometry, medicine, marketing, among others, there are an interest in studying latent variables. Such variables can not be measured directly. Some examples of latent variables are: the knowledge in some cognitive field, depression level or genetic predisposition in manifesting some disease. Generally, these variables are inferred based on responses of subjects to certain measurement instruments, which are composed by items (questions), for example, a mathematical test, a quality of life questionnaire or a survey about characteristics of some product. The item response theory (IRT) is a set of models that allows to make inference about latent variables, based on the subject's performance on measurement instruments. Specifically, such models represent the probability of subjects get some score in each item, taking into account the subject's latent trait and certain test's characteristics (and eventually other additional information).

Several aspects of the IRT have been explored. In fact, the development of more flexible models, suitable estimation methods and model fit assessment tools, have been the main focus of several works in the literature. Important discussions about IRT in their different aspects can be find in Hambleton and Swaminathan (1985), Hambleton et al. (1991), Lord (1980), Lord and Novick (1968) e Andrade et al. (2000).

1.2 Motivation and literature review

Due the application of the IRT in several research fields and also due to the appearance of more complex studies, suitable IRT models have been developed in order to describe properly the characteristics of these studies. Therefore, many approaches have been studied as: Skew models, see Azevedo et al. (2011), Santos et al. (2013) and Bazán et al. (2004), that consider skewed latent trait distributions and/or skewed link function. Multidimensional models, see Nojosa (2001), which assume more than one latent trait to model the probability responses. Longitudinal models, see Azevedo (2003), Azevedo et al. (2016), Tavares (2001), Andrade and Tavares (2005) and Tavares and Andrade (2006b) which modeling between latent traits dependence. These and other class of IRT models allow the IRT application in a wide range of situations and also lead to more accurate results.

The focus of this work is the longitudinal IRT models. Longitudinal data occur when experimental units are followed along different measurement occasions (or time-points) that

can not be mutually randomized. Along the measurement occasions, we measured from these experimental units, characteristics of interest. In the IRT context, we observed the response of subjects to items belonging to some measurement instrument (such that, cognitive tests, psychiatric questionnaires, educational tests, among others), along the different occasions (as scholar grades) that can be partially or totally different to each other, such that, time-specific measurements are nested within respondents. Due this nested structure, some within-subject dependence is expected, see Andrade and Tavares (2005), Azevedo et al. (2012b). The correct characterization of this dependence is fundamental to obtain trustable results on longitudinal IRT data analysis.

Some longitudinal IRT 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 (2006a) developed an IRT model to estimate ability distribution parameters, considering several covariance structures when item parameters are known and unknown, respectively. Recently, Azevedo et al. (2016) proposed a general class of longitudinal IRT models with multivariate normal distributions for the latent traits, considering a Bayesian framework. This class takes into account important features of the longitudinal data, as heteroscedasticity of population variances and serial correlation. Furthermore, the authors proposed some Bayesian model fit assessment tools. However, the multivariate distributions considered to simulate latent traits are computationally very expensive, specially when the number of subjects and the number of measurement occasions are large. In addition, even the authors explored some structured covariance matrix, their approach to estimate covariance structures involves the estimation of a non structured full matrix. It means that, despite to the restricted correlation pattern, the number of parameters to be estimated are not reduced and it increase as the number of measurement occasions increase. In this sense, it can also be difficult to handle unbalanced data, caused by dropouts and/or inclusion of persons during the study.

Additionally, a common assumption in the previous IRT longitudinal models is the symmetric multivariate normality of the latent traits distributions. This assumption is often unrealistic and can lead to biased estimates, see Azevedo et al. (2011) and Azevedo et al. (2012a) and references therein. In longitudinal studies, it is very common to observe asymmetric behavior of the latent traits distributions, since that the inclusion/exclusion of subjects along the study may induce different behaviors for these distributions. Moreover, from an educational point of view, it is expected to observe a growing in the latent traits of the subjects, resulting in negative asymmetry on the latent traits distribution.

Concerning to the responses modeling, most of the previous works did not consider the effect of guessing on the probability of response in multiple choice items. This effect is important to represent properly the latent traits with subjects with very low probability of right score the item. Therefore, we consider the three-parameters probit model, Baker and Kim (2004).

Therefore, our proposal consists on a three parameter probit model (Baker and Kim, 2004) and for the latent trait structure, we will basically consider two approaches: antedependence models and copula modeling.

Antedependence models have been applied in the longitudinal data analysis in several works, see Pourahmadi (1999), Cepeda-Cuervo and Núñez-Antón (2009), Daniels and Pourahmadi (2002) and Nunez-Anton and Zimmerman (2000), for example. These are autoregressive models based on a modified Cholesky decomposition of the precision matrix. It allows a parsimonious

modeling of the within-subject covariance structure while heeding its positive-definiteness, that is of great importance in longitudinal data analysis. Also, this approach is quite flexible and allows to handle multivariate distributions through the univariate conditional distributions. This feature can reduce computational cost in MCMC algorithms, compared to the multivariate approach. It also allows to represent properly a wide range of specific correlation patterns and latent trait distributions, without to consider additional random effects for the model. Furthermore, such modeling induce a conditional univariate structure in the latent trait distributions, which is quite interesting to handle unbalanced data.

Another proposal to handle the latent traits structure is the copula modeling. Copula models have become one of the most widely used tools in the modeling of 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 among the variables is captured using an 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 to represent marginal 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 this work we will explore the copula approach in the modeling of IRT longitudinal data.

An important stage on the IRT approach application concerns to the estimation process (Baker and Kim, 2004, Azevedo, 2003). Particularly, the IRT models requires estimation methods that take into account their complexity. In the last years, the Bayesian approach have been successfully applied. see Gilks et al. (1996), Albert (1992), Azevedo et al. (2011), for example. The Monte Carlo Markov Chains (MCMC) algorithms allows to consider different latent traits distributions easily, selection of covariance structures, missing responses caused by design, missing at random, among other characteristics. This algorithms provide empirical approximation for the posterior marginal distributions by considering the so called full conditional distributions, see Gamerman and Lopes (2006). Basically, for full conditional distributions with know form, the Gibbs sampling is considered, otherwise, auxiliary algorithms as Metropolis-Hastings should be adopted. Therefore, the development of suitable MCMC algorithms allows a broader application of the IRT models.

According to the longitudinal IRT structure the within-subject latent traits are expected to be correlated. In this case, the MCMC algorithms tend to generate high autocorrelated chains, see Gamerman and Lopes (2006). In order to reduce autocorrelation, we consider the so called Forward Filtering Backward Sampling (FFBS), see Carter and Kohn (1994) and Frühwirth-Schnatter (1994). The key idea of this sampler is to sample the correlated parameters in blocks, considering the dependence structure of the latent traits in the sampling process.

Another relevant aspect in the estimation process of the longitudinal IRT models is the estimation of the correlation parameters. As we will see in the next chapters, when structured covariance matrices are used, auxiliary algorithms need to be considered since the full conditional distributions of the correlation parameters have unknown forms. A common choice in this situation is the Metropolis-Hastings algorithm. Such algorithm requires a proposal distribution (a common choice is a random walk centered on the previous result) to generate proposed values. This value is accepted or rejected according to a transition probability. Then, the acceptance 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 proposed by Murray et al. (2012). The so called *Single-Variable Exchange* algorithm (SVE) differ in the way to propose *candidates* values. The efficiency of the SVE algorithm can be improved by considering the oversampling procedure proposed by Marsman (2014). The key idea of this strategy, is to concentrate more probability mass on transition kernels with high acceptance probability. In this sense, a suitable number of *i.i.d* proposal values are simulated, instead of a single one. The most likely value (in terms of likelihood) is taken.

A third aspect concerns to the model fit assessment. For more complex IRT models, as the longitudinal IRT models, there is still few proposals in this sense. In this work posterior predictive checking mechanisms studied by Azevedo et al. (2012b) and Santos et al. (2013) are considered. It is based on the predictive distribution of some discrepancy measures, and allows to check general and specific aspects (assumptions) of the model.

Another situation of interest, in the context of the IRT longitudinal data analysis, occur when the subjects are nested into different different groups. For example: the assessment of students from public and private schools followed along grades concerning some knowledge field, the measuring of quality of life of men and women along some weeks, the measuring of psychiatric condition of male and female patients along some years, and so on. This data structure was named by Azevedo et al. (2015) as Longitudinal Multiple Group Item Response Data. In this kind of IRT data the group heterogeneity can reflect different behaviors, as well as, the longitudinal within group structure can induce a correlation pattern between the measures of the same subject. The multiple group longitudinal IRT model proposed by Azevedo et al. (2015) unifies two methodologies. The multiple group IRT model (Bock and Zimowski, 1997), that allows a simultaneously equating estimation process which leads to more accurate results than a posterior equating, see Kolen and Brennan (2004) and the longitudinal IRT model (Azevedo et al., 2012b), that takes into account the correlation structure of the latent traits. However, this methodology does not consider the unbalanced data design caused by dropouts and/or inclusion of subjects, that is very common in longitudinal studies. Also the asymmetry of the latent trait distributions are not considered. Therefore, we will extend the work of Azevedo et al. (2012b) to allow asymmetry on the latent trait distributions and also unbalanced designs.

Another feature that we will consider in our methodology, is the modeling of the mean structure through specific parametric growth curves. The growth curves modeling can provides a considerable parameters reduction especially in the case of many measurement points.

1.3 Objectives and outline

1.3.1 Objectives

The main objective of this work is to develop classes of models for analyzing IRT longitudinal data considering usual normal and skew normal latent trait distributions, multiple group and unbalanced data structures and also some growth curve models for the mean of the latent trait distributions. Such models will extend the works of Azevedo et al. (2011) and Santos et al. (2013), concerning the longitudinal structure incorporation and the work of Azevedo et al. (2015), Tavares and Andrade (2006b) and Tavares and Andrade (2006a) concerning the skewed latent trait distributions and the unbalanced data modeling. Our specific goals are:

- Develop a three parameter probit model to analyze longitudinal IRT data with both normal and skew-normal latent trait distributions;
- Develop a flexible approach that allows to consider different covariance structures for the latent traits considering both antedependence and copula modeling under a Bayesian framework;
- Propose suitable MCMC algorithms for model parameters estimation;
- Adapt some posterior predictive techniques for model fit assessment;
- Consider a multiple group structure and growth curves for mean of the latent traits;
- Perform simulation studies to assess the efficiency of our models and estimation algorithms and to compare competitors methods;
- Apply our methodologies in real data sets.

1.3.2 Outline

The next four chapters in this thesis are written in an article style, intended to be self-contained, and some overlap could not be avoided.

In Chapter 2 we discuss a Cholesky decomposition based modeling for single-group longitudinal IRT data considering symmetric normal latent trait distributions. In Chapter 3, the methodology developed in Chapter 2 is extended to consider skewed latent trait distributions. In Chapter 4 a multiple-group longitudinal IRT model with skewed latent trait distributions and growth curves structures for the latent traits is proposed. In Chapter 5 a longitudinal IRT model with skewed latent trait distributions, considering Gaussian copulas is developed, as an alternative for the model presented in Chapter 2. Finally, in Chapter 6 we present the final conclusions.

Chapter 2

Bayesian general Cholesky decomposition based modeling of longitudinal IRT data.

Abstract

In this chapter 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 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 considering regression (growth curves) and multilevel structures for the latent traits, among other advantages. In this chapter 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 which have some structure of common items. Using an appropriate augmented data structure, a longitudinal IRT model is developed through the Pourahmadi's approach. The parameter estimation, model fit assessment and model comparison were implemented through a hybrid MCMC algorithm, such that when the full conditionals are not known, the SVE (Single Variable Exchange) 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.

2.1 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 of 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 is, the time-specific measurements within-subjects, it is expected to observe some within-subject dependence. Within 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 will be observed for this quantities, see Andrade and Tavares (2005), Azevedo et al. (2012b).

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 (2006a) 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 chapter is outlined as follows. In Section 2.2, the longitudinal IRT antedependence model is presented along with some of its properties. In Section 2.3, we describe all steps of our MCMC algorithm. In Section 2.4 some simulation studies are presented to study the efficiency of our model and MCMC algorithm, concerning some features of interest. In Section 2.5 a real data from the Brazilian school development program is analyzed and some model fit assessment are presented. Finally, in Section 2.6 we presented some conclusions and remarks.

2.2 Modeling

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: θ_{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 , $\mathbf{Y}_{.jt} = (Y_{1jt}, \dots, Y_{I_tjt})'$ is the response vector of subject j at the time-point t , $\mathbf{Y}_{..} = (\mathbf{Y}'_{..1}, \dots, \mathbf{Y}'_{..T})'$ is the entire response matrix and $(y_{ijt}, \mathbf{y}'_{.jt}, \mathbf{y}'_{..})'$ 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.

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

$$Y_{ijt} | \theta_{jt}, \boldsymbol{\zeta}_i \sim \text{Bernoulli}(P_{ijt}), \quad (2.2.1)$$

$$P_{ijt} = \mathbb{P}(Y_{ijt} | \theta_{jt}, \boldsymbol{\zeta}_i) = c_i + (1 - c_i) \Phi(a_i \theta_{jt} - b_i). \quad (2.2.2)$$

$$\boldsymbol{\theta}_j | \boldsymbol{\eta}_\theta \sim D_T(\boldsymbol{\eta}_\theta), \quad (2.2.3)$$

where $D(\cdot)$ stands for some T -variate distribution indexed by the parameters $\boldsymbol{\eta}_\theta$. In equation

(2.2.2), 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 correlation pattern identification is very importante to explain the growth in latent traits, as pointed out by Azevedo et al. (2016). In this work, we will adapt for IRT context, the general procedure of covariance matrix estimation proposed by Pourahmadi (1999). Such approach is based on the Cholesky decomposition of the covariance matrix inverse and it allows to represent a wide range of the variance-covariance structures.

2.2.1 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 $\text{Cov}(\boldsymbol{\theta}_j) = \boldsymbol{\Sigma}_\theta$.

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

$$\theta_{jt} = \mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}) + \varepsilon_{jt}, \quad t = 1, 2, \dots, T, \quad (2.2.4)$$

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

In matrix form, we have:

$$\boldsymbol{\varepsilon}_j = \mathbf{L}(\boldsymbol{\theta}_j - \boldsymbol{\mu}_\theta) \quad \Leftrightarrow \quad \boldsymbol{\theta}_j = \boldsymbol{\mu}_\theta + \mathbf{L}^{-1}\boldsymbol{\varepsilon}_j. \quad (2.2.5)$$

This model was named by Zimmerman and Núñez-Antón (2009) unstructured antedependence model. The random vector $\boldsymbol{\varepsilon}_j = (\varepsilon_{j1}, \varepsilon_{j2}, \dots, \varepsilon_{jT})'$ are uncorrelated with $\text{Cov}(\boldsymbol{\varepsilon}_j) = \mathbf{D}$, where \mathbf{D} is a diagonal matrix $\text{diag}(d_1, d_2, \dots, d_T)$ 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 (2.2.5) and using the definition of \mathbf{D} we have that,

$$\text{Cov}(\boldsymbol{\varepsilon}_j) = \mathbf{L}\text{Cov}(\boldsymbol{\theta}_j - \boldsymbol{\mu}_\theta)\mathbf{L}' = \mathbf{L}\boldsymbol{\Sigma}_\theta\mathbf{L}' = \mathbf{D}. \quad (2.2.6)$$

Therefore the matrix \mathbf{L} diagonalize 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 can 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:

$$\Sigma_{\theta} = \begin{pmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta} & \sigma_{\theta_1}\sigma_{\theta_3}\rho_{\theta}^2 \\ \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta} & \sigma_{\theta_2}^2 & \sigma_{\theta_2}\sigma_{\theta_3}\rho_{\theta} \\ \sigma_{\theta_1}\sigma_{\theta_3}\rho_{\theta}^2 & \sigma_{\theta_2}\sigma_{\theta_3}\rho_{\theta} & \sigma_{\theta_3}^2 \end{pmatrix}.$$

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 by using equation (2.2.5) we can obtain the following linear model:

$$\begin{aligned} \theta_{j1} - \mu_{\theta_1} &= \varepsilon_{j1}, \\ \theta_{jt} - \mu_{\theta_t} &= \frac{\sigma_{\theta_t}}{\sigma_{\theta_{t-1}}}\rho_{\theta}(\theta_{j(t-1)} - \mu_{\theta_{t-1}}) + \varepsilon_{jt}, \quad t = 2, \dots, T. \end{aligned} \quad (2.2.7)$$

Note that the parameters $(d_1, d_2, \dots, d_T, \phi_{21}, \phi_{31}, \phi_{32}, \dots, \phi_{T(T-1)})'$ are an one-to-one mapping of the parameters of interest $(\sigma_{\theta_1}^2, \sigma_{\theta_2}^2, \dots, \sigma_{\theta_T}^2, \rho_{\theta})'$. 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 the expression (2.2.7)). In such cases the Cholesky decomposition is obtained numerically. Table 2.1 presents some examples of structured matrices which are considered in this work.

Table 2.1: Structured covariance matrices used in this work.

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

In order to represent the multivariate structure of the latent traits (see equation 2.2.3) we will consider the following antedependence model:

$$\theta_{jt} = \mu_t + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_k) + \varepsilon_{jt}, \quad \varepsilon_{jt} \stackrel{i.i.d.}{\sim} N(0, d_t), \quad t = 1, 2, \dots, T. \quad (2.2.8)$$

Equivalently,

$$\varepsilon_j = \mathbf{L}(\boldsymbol{\theta}_j - \boldsymbol{\mu}_\theta), \quad \varepsilon_j \stackrel{ind.}{\sim} N_T(\mathbf{0}, \mathbf{D}), \quad (2.2.9)$$

where \mathbf{D} is a diagonal matrix described before. By the normal distribution properties and the property (2.2.6) we can see that:

$$\boldsymbol{\theta}_j \sim N_T(\boldsymbol{\mu}_\theta, \mathbf{L}^{-1}\mathbf{D}(\mathbf{L}^{-1})') \equiv N_T(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta). \quad (2.2.10)$$

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

$$\theta_{jt} \sim N(\mu_{\theta_t}, \sigma_{\theta_t}^2), \quad t = 1, 2, \dots, T, \quad (2.2.11)$$

where

$$\begin{aligned} \sigma_{\theta_1}^2 &= d_1 \\ \sigma_{\theta_t}^2 &= d_t + \sum_{k=1}^{t-1} l_{tk}^2 d_k, \end{aligned} \quad (2.2.12)$$

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

2.3 Bayesian Estimation and MCMC Algorithms

In order to make easier the implementation of the MCMC algorithms we used the augmented data approach to represent our IRT model, see Tanner and Wong (1987). For the three-parameter models we can use the augmented data scheme 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} given $Y_{ijt} = y_{ijt}$ is given by

$$\begin{aligned} \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 1, \theta_{jt}, \zeta_i) &\propto \Phi(a_i \theta_{jt} - b_i) \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 1, \theta_{jt}, \zeta_i) &\propto c_i (1 - \Phi(a_i \theta_{jt} - b_i)) \\ \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 0, \theta_{jt}, \zeta_i) &= 0 \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 0, \theta_{jt}, \zeta_i) &= 1. \end{aligned} \quad (2.3.1)$$

Therefore the augmented variables $\mathbf{Z} = (Z_{111}, \dots, Z_{1n_1 1}, \dots, Z_{In_T T})'$, are given by

$$Z_{ijt} | (\theta_{jt}, \zeta_i, w_{ijt}) = \begin{cases} N(a_i \theta_{jt} - b_i, 1) \mathbb{I}_{(z_{ijt} \geq 0)}, & \text{if } w_{ijt} = 1, \\ N(a_i \theta_{jt} - b_i, 1) \mathbb{I}_{(z_{ijt} < 0)}, & \text{if } w_{ijt} = 0. \end{cases} \quad (2.3.2)$$

The original response can be represented by

$$Y_{ijt} = \mathbb{I}(Z_{ijt} \geq 0)\mathbb{I}(W_{ijt} = 1) + \mathbb{I}(Z_{ijt} < 0)\mathbb{I}(W_{ijt} = 0), \quad (2.3.3)$$

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

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

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 assume 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

$$\begin{aligned} L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta} | \mathbf{z}_{..}, \mathbf{w}_{..}, \mathbf{y}_{..}, \mathbf{I}) &\propto \prod_{t=1}^T \prod_{j=1}^{n_t} \prod_{i \in I_{jt}} \exp \left\{ -.5(z_{ijt} - a_i \theta_{jt} + b_i)^2 \right\} \mathbb{I}_{(z_{ijt}, w_{ijt})} \\ &\times p(w_{ijt} | y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i), \end{aligned} \quad (2.3.4)$$

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-point t and $p(w_{ijt} | y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i)$ follows from (2.3.1).

2.3.1 Prior and posterior distributions

The joint prior distribution is assumed to be as

$$p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta} | \boldsymbol{\eta}_{\zeta}, \boldsymbol{\eta}_{\eta}) = \left\{ p(\theta_{j1} | \boldsymbol{\eta}_{\theta_1}) \prod_{t=2}^T \prod_{j=1}^{n_t} p(\theta_{jt} | \boldsymbol{\theta}_{j(1:t-1)}, \boldsymbol{\eta}_{\theta_t}) \right\} \left\{ \prod_{i=1}^I p(\boldsymbol{\zeta}_i | \boldsymbol{\eta}_{\zeta}) \right\} \left\{ \prod_{t=1}^T p(\boldsymbol{\eta}_{\theta_t} | \boldsymbol{\eta}_{\eta}) \right\}, \quad (2.3.5)$$

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 (2.2.8). For the item parameters we have:

$$p(\boldsymbol{\zeta}_{i(-c_i)}) \propto \exp \left[-.5(\boldsymbol{\zeta}_{i(-c_i)} - \boldsymbol{\mu}_{\zeta})' \boldsymbol{\Psi}_{\zeta}^{-1} (\boldsymbol{\zeta}_{i(-c_i)} - \boldsymbol{\mu}_{\zeta}) \right] \mathbb{I}_{(a_i > 0)}, \quad (2.3.6)$$

and

$$c_i \sim \text{Beta}(a_c, b_c), \quad (2.3.7)$$

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

$$\begin{aligned}
\mu_t &\sim N(m_\mu, \sigma_\mu^2), \\
d_t &\sim IG(a_d, b_d), \quad t = 1, \dots, T, \\
\phi_{tk} &\sim N(\mu_\phi, \sigma_\phi^2) \quad t = 2, \dots, T \text{ and } k = 1, \dots, t-1,
\end{aligned} \tag{2.3.8}$$

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 2.1) the prior distributions for correlation parameters are directly specified as:

$$\rho_{\theta_t} \stackrel{i.i.d}{\sim} N(\mu_\rho, \sigma_\rho^2) \mathbb{I}_{[0,1]}, \quad t = 1, 2, \dots, T-1, \tag{2.3.9}$$

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 (2.3.4) and the prior distribution in equations (2.2.8), (2.3.6), (2.3.7) and (2.3.8), the joint posterior distribution is given by:

$$\begin{aligned}
p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, |\mathbf{z}_{..}, \mathbf{w}_{..}, \mathbf{y}_{..}) &\propto \left\{ \prod_{t=1}^T \prod_{j=1}^n \prod_{i \in I_{jt}} \exp \left\{ -0.5(z_{ijt} - a_i \theta_{jt} + b_i)^2 \right\} p(w_{ijt} | y_{ijt}, \theta_{jt}, \boldsymbol{\zeta}_i) \mathbb{I}_{(z_{ijt}, w_{ijt})} \right\} \\
&\times \left\{ \prod_{j=1}^{n_t} p(\theta_{j1}, \boldsymbol{\eta}_{\theta_t}) \prod_{t=2}^T p(\theta_{jt} | \theta_{j(1:t-1)}, \boldsymbol{\eta}_{\theta_t}) \right\} \left\{ \prod_{i=1}^I p(\boldsymbol{\zeta}_i | \boldsymbol{\eta}_\zeta) \right\} \\
&\times \left\{ \prod_{t=1}^T p(\boldsymbol{\eta}_{\theta_t} | \boldsymbol{\eta}_\eta) \right\},
\end{aligned} \tag{2.3.10}$$

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. 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:

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

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

and

$$B_t = \begin{cases} \left(\sum_{i \in I_{jt}} a_i^2 + \frac{1}{d_t} + \frac{\phi_{(t+1)t}^2}{d_{t+1}} \right)^{-1}, & t = 1, \dots, T-1 \\ \left(\sum_{i \in I_{jt}} a_i^2 + \frac{1}{d_t} \right)^{-1}, & t = T. \end{cases} \tag{2.3.11}$$

For the items parameters we have:

$$\begin{aligned} \zeta_{i(-c_i)} | (\cdot) &\sim N(\widehat{\Psi}_{\zeta_i} \widehat{\zeta}_i, \widehat{\Psi}_{\zeta_i}), \\ \widehat{\zeta}_i &= (\Theta_{i.})' \mathbf{z}_i + \Psi_{\zeta}^{-1} \boldsymbol{\mu}_{\zeta}, \\ \widehat{\Psi}_{\zeta_i} &= \left[(\Theta_{i.})' (\Theta_{i.}) + \Psi_{\zeta}^{-1} \right]^{-1}, \\ \Theta_{i.} &= [\boldsymbol{\theta} - \mathbf{1}_n] \bullet \mathbb{1}_i, \end{aligned} \quad (2.3.12)$$

(\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 guessing parameters,

$$c_i | (\cdot) \sim \text{Beta}(s_i + a_c - 1, t_i - s_i + b_c - 1), \quad (2.3.13)$$

where

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

For the population parameters we have the following full conditional distributions:

$$\mu_t | (\cdot) \sim N(A_t a_t, A_t), \quad (2.3.14)$$

where

$$\begin{aligned} a_t &= \frac{1}{d_t} \sum_{j=1}^{n_t} \left[\theta_{jt} - \sum_{k=1}^{t-1} \phi_{tk} (\theta_{jk} - \mu_k) \right] + \frac{m_{\mu}}{\sigma_{\mu}^2} \text{ and } A_t = \left(\frac{n_t}{d_t} + \frac{1}{\sigma_{\mu}^2} \right)^{-1}. \\ d_t | (\cdot) &\sim IG\left(\frac{n_t}{2} + a_d, \frac{s_t}{2} + b_d\right) \text{ where,} \end{aligned} \quad (2.3.15)$$

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

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

where

$$Q_{tk} = \left(\frac{\sum_{j=1}^{n_t} (\theta_{j(t-1)} - \mu_{\theta_{t-1}})^2}{d_t} + \frac{1}{\sigma_{\phi}^2} \right)^{-1} \quad (2.3.16)$$

$$q_{tk} = \frac{1}{d_t} \sum_{j=1}^{n_t} (\theta_{jk} - \mu_{\theta_k}) (\theta_{jt} - \mu_{\theta_t} - \sum_{k \neq t} \phi_{tk} (\theta_{jk} - \mu_{\theta_k})).$$

for all $t = 2, \dots, T$ and $k = 1, \dots, (t-1)$. The marginal variances can be recovered by using the relation (2.2.12). Therefore, a full Gibbs sampling (FGS) to the model parameters estimation is provided by Algorithm 2.1.

Algorithm 2.1 Full Gibbs sampling algorithm

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

The algorithm 2.1 is suitable to fit unstructured covariance matrices. For the structured matrices (see Table 2.1), we need to use an auxiliary algorithm to sampling correlation 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 wich can handle these two issues.

2.3.2 An alternative 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 2.2.8). A common way to sample from the joint full conditional distribution of θ_j . it is to consider univariate full conditional distributions, in order to sample a time-specific latent trait given all other ones, through Gibbs sampling. It means that: $\theta_{jt}|\theta_{j(-t)}$ for all $t = 1, \dots, T$, where $\theta_{j(-t)}$ is the latent traits vector without the t -th component. This is the idea of the full conditional distribution presented in equation (2.3.11). 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 **Forward Filtering Backward Sampling** (FFBS). In the following we presented this approach adapted to our problem.

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:

$$u_t = F_t^\top \theta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \quad (2.3.17)$$

$$\theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, \Omega_t), \quad (2.3.18)$$

where u_t is a sequence of observations at the time, conditionally independent given θ_t and σ_t^2 . 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 model:

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

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

FFBS Algorithm

The FFBS algorithm basically consist 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 $\theta_{jt}|\theta_{j(t-1)} \sim N(\mu_t + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_k), d_t)$. By properties of the normal distribution, these specifications can be combined leading to the marginal distribution:

$$\theta_{jt}|\mathbf{z}_j^{t-1} \sim N(a_{jt}, R_{jt}), \quad (2.3.19)$$

where

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

Thus,

$$\theta_{jt}|\mathbf{z}_j^t \sim N(m_{jt}, C_{jt}), \quad (2.3.20)$$

where

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

Equation (3.4.26) is referred in the literature as *Kalman Filter*. Therefore, the backward distributions are given by:

$$\theta_{jt}|\theta_{j(t+1)}, \mathbf{z}_j^t \sim N(m_{\theta_t}, C_{\theta_t}), \quad (2.3.21)$$

where

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

where $\alpha_{t+1} = \mu_{t+1} - \phi_{t+1,t}\mu_t + \sum_{k=1}^{t-1} \phi_{t+1,t}(\theta_{jk} - \mu_k)$, for all $j = 1, \dots, n_t$ and $t = 1, \dots, T$. Then, a scheme to sample from the full conditional distribution of $\boldsymbol{\theta}_j$ is presented in algorithm 2.2.

Algorithm 2.2 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 $t = T$.

2.3.3 Oversampling in the Single-Variable Exchange algorithm

In Section 2.3.1 was presented full Gibbs sampling algorithm that is useful for the 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 of the structured matrices.

An usual way for sampling from an intractable posterior distribution it is to use the *Metropolis-Hastings* algorithm, see Gamerman and Lopes (2006). It requires a proposal distribution (a common choice is a random walk centered on the previous result) to generate proposed values. This value is accepted or rejected according to a transition probability. Then, the acceptance rate of the proposed values will depends 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 proposed by Murray et al. (2012), see algorithm (2.3). The so called *Single-Variable Exchange* algorithm (SVE) differs in the way to propose *candidates* values.

Specifically, consider the following posterior distribution:

$$p(\vartheta|\mathbf{x}) \propto p(\mathbf{x}|\vartheta)p(\vartheta), \quad (2.3.22)$$

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 (2.3.22) is given by algorithm 2.3.

Algorithm 2.3 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** ($u < \pi(\vartheta^{(m-1)} \rightarrow \vartheta^{(m)})$) **then**
 - 5: $\vartheta^{(m-1)} = \vartheta^{(m)}$
 - 6: **end if**
-

where,

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

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

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 to simulate a number of i.i.d proposal values, instead of simulate a single one. Let $t(\mathbf{x})$ a sufficient statistic for the parameter ϑ and $t(\mathbf{x}^{(m)})$ the corresponding sufficient statistic of the proposed value $\vartheta^{(m)}$. Let s being the number of proposed values each one with its own sufficient statistic, we have to choose the one whose statistic $t(\mathbf{x}^{(m)})$ is closest to the observed $t(\mathbf{x})$, see algorithm (2.4).

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

Algorithm 2.4 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** $(u < \pi(\vartheta^{(m-1)} \rightarrow \vartheta^{(m)}))$ **then**
 - 9: $\vartheta^{(m-1)} = \vartheta^{(m)}$
 - 10: **end if**
-

as defined in Section 2.2 and $p(\boldsymbol{\theta}_{..}|\boldsymbol{\mu}_\theta, \boldsymbol{\phi}, \mathbf{d})$ denoting the likelihood generated by the antedependence model defined in equation (2.2.8), that is

$$\begin{aligned}
 p(\boldsymbol{\theta}_{..}|\boldsymbol{\mu}_\theta, \boldsymbol{\phi}, \mathbf{d}) &= \prod_{j=1}^{n_t} p(\theta_{j1}|\mu_{\theta_1}, d_1) \prod_{t=2}^T p(\theta_{jt}|\mu_{\theta_t}, \boldsymbol{\phi}_t, d_t) \\
 &\propto \prod_{j=1}^{n_t} \frac{1}{\sqrt{d_1}} \exp \left\{ -\frac{1}{2d_1} (\theta_{j1} - \mu_{\theta_1})^2 \right\} \\
 &\times \prod_{t=1}^T \frac{1}{\sqrt{d_t}} \exp \left\{ -\frac{1}{2d_t} \left[(\theta_{jt} - \mu_{\theta_t}) - \sum_{k=1}^{t-1} \phi_{tk} (\theta_{jk} - \mu_{\theta_k}) \right]^2 \right\}, \quad (2.3.23)
 \end{aligned}$$

where $\boldsymbol{\mu}_\theta = (\mu_{\theta_1}, \dots, \mu_{\theta_T})'$, $\boldsymbol{\phi} = (\phi_{21}, \phi_{31}, \phi_{32}, \dots, \phi_{T(T-1)})'$, $\mathbf{d} = (d_1, \dots, d_T)'$ and $\boldsymbol{\phi}_t$ denotes the elements correspondents to the time-point t . Algorithm 2.5 present the correlation parameters sampler for the ARH(1) model. In this example the sufficient statistic is the first-order sample correlation.

Algorithm 2.5 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 $\boldsymbol{\theta}_{\dots s}^{(m)}$ from the model (2.2.8)

4: Compute $r_1(\boldsymbol{\theta}_{\dots s}^{(m)})$ the first-order sample correlation

5: **end for**

6: Choose the $\rho_{\theta_s}^{(m)}$ and $\boldsymbol{\theta}_{\dots s}^{(m)}$ whose $r_1(\boldsymbol{\theta}_{\dots s}^{(m)})$ is closest to $r_1(\boldsymbol{\theta}_{\dots}^{(m-1)})$

7: Set $\rho_\theta^{(m)} = \rho_{\theta_s}^{(m)}$ and $\boldsymbol{\theta}_{\dots}^{(m)} = \boldsymbol{\theta}_{\dots s}^{(m)}$ the candidate values

8: Build the ARH(1) proposed matrix $\Sigma_{\rho_\theta}^{(m)}$ using `AR1.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**

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

then

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

13: **end if**

When the structured matrix present two or more correlation parameters, the algorithm 2.5 can be applied independently for each one 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 2.5 to allow accept/reject proposed values jointly. More details about the SVE are found in Appendix A.

In summary, we can combine the full Gibbs sampling with the FFBS algorithm by replacing step 4 of the algorithm 2.1 by algorithm 2.2, 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 2.1 by a SVE procedure, in order to simulate correlation parameters of a structured matrix. The algorithms 2.6 and 2.7 summarized these options.

Algorithm 2.6 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 $W_{ijt}|\cdot$ for all $i = 1, \dots, I, j = 1, \dots, n_t$ e $t = 1, \dots, T$.
 - 3: Simulate Z_{ijt} from $Z_{ijt}|\cdot$ for all $i = 1, \dots, I, j = 1, \dots, n_t$ e $t = 1, \dots, T$.
 - 4: Simulate θ_{jt} using algorithm 2.2 for all $j = 1, \dots, n_t$ e $t = 1, \dots, T$.
 - 5: Simulate ζ_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 μ_{θ_t} from $\mu_{\theta_t}|\cdot$ for all $t = 1, \dots, T$.
 - 8: Simulate d_t from $d_t|\cdot$ for all $t = 1, \dots, T$.
 - 9: Simulate ϕ_{tk} from $\phi_{tk}|\cdot$ for all $t = 2, \dots, T$ and $k = 1, \dots, t - 1$.
-

Algorithm 2.7 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 $W_{ijt}|\cdot$ for all $i = 1, \dots, I, j = 1, \dots, n_t$ e $t = 1, \dots, T$.
 - 3: Simulate Z_{ijt} from $Z_{ijt}|\cdot$ for all $i = 1, \dots, I, j = 1, \dots, n_t$ e $t = 1, \dots, T$.
 - 4: Simulate θ_{jt} using algorithm 2.2 for all $j = 1, \dots, n_t$ e $t = 1, \dots, T$.
 - 5: Simulate ζ_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 μ_{θ_t} from $\mu_{\theta_t}|\cdot$ for all $t = 1, \dots, T$.
 - 8: Simulate d_t from $d_t|\cdot$ for all $t = 1, \dots, T$.
 - 9: Simulate ρ_{θ_t} using a SVE procedure for all necessary correlation parameters.
-

2.4 Simulation Studies

To study the performance of the FFBS procedure, we compared algorithms 2.1 and 2.6. 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 of the sample. 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 study we consider a real data concerning to the Brazilian school development study (see, Section 2.5). The model 2.2.2 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 2.2 presents the average value for each sampler. We can see that the FFBS outperforms the FGS algorithm, considering both ESS and ESS/h.

Table 2.2: Effective sample size and effective sample size per hour considering two latent traits samplers

	ESS	ESS/h
FFBS	7045.69	286.90
FGS	6753.43	275.56

Henceforth, we will only consider algorithms 2.6 and 2.7 for unstructured and structured matrices, respectively.

2.4.1 Parameter Recovery

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 2.1. Responses of $n_t = 1500$ subjects, along $T = 6$ time-points were simulated according to the model (2.2.8), for each covariance matrix of Table 2.1. The items 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) (see Table B.1 in appendix). 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 (2.2.8) 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 is considered as the reference one. Then we fixed increasing population means on the (0, 1) scale (respective the mean and the variance of the reference time-point), implying a growing for the latent traits. This is an expected behavior in educational longitudinal studies, for example, see Santos et al. (2013) and Azevedo et al. (2012b). The population variances induce a decreasing and then an increasing in the time-heterogeneity. Similarly to the the real data analyzed in next section, we fixed from high to moderate values for population correlation in order to have a similar pattern to that observed in the real data. The correlation values can be seen in Table 2.3.

Table 2.3: 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

Table 2.4 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 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 concentration toward zero. The discrimination parameters are assumed to vary reasonably around a satisfactory discrimination power 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.

Table 2.4: 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, see Gelman et al. (2004), were monitored. We generate three chains based on three different sets of starting values. The inspection of them 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: $\text{cor}(\vartheta_l, \hat{\vartheta}_l)$, 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: $\sqrt{\frac{1}{n_p} \sum_{l=1}^{n_p} (\text{MVAR} + (\vartheta_l - \hat{\vartheta}_l)^2)}$, with $l = 1, \dots, n_p$ where n_p denotes the number of parameters.

Tables 2.5 to 2.7 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% credibility intervals cover the respective true value. 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 2.8 presents the results for the latent traits and item parameters. They indicate that our algorithm recovered all parameters properly. For the guessing parameter correlations are not useful, since the true values have small variability compared to the estimates. Therefore, they were not presented. Figure 2.1 presents 95% 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 2.5: Results for the population means parameters

		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 2.6: Results for the population variance parameters

		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 2.7: Results for the population correlation parameters.

		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 2.8: 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	-	-.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	-	.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	-	.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	-	.003	.025	.003	.062

Unbalanced data

For the next example we simulated an unbalanced data design, representing a longitudinal study with dropouts. Therefore, we simulate data considering the same scenario described in Subsection 2.4.1 but, for the sake of simplicity, we consider only the ARH(1) covariance matrix. The number of subjects per time were $n_1 = 1500, n_2 = 1400, n_3 = 1400, n_4 = 1400, n_5 = 1350, n_6 = 1350$. Tables 2.9 and 2.10 present the results of the parameter recovery. 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 2.1 and 2.2 present the estimates of the item parameters for the balanced and unbalanced case, respectively. 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 2.9: Results for the population parameters for the unbalanced data study

	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]
$\sigma_{\theta_1}^2$	1.000	1.000	–	–
$\sigma_{\theta_2}^2$.770	1.491	.196	[1.185, 1.960]
$\sigma_{\theta_3}^2$.300	.472	.053	[.375, .576]
$\sigma_{\theta_4}^2$.380	.655	.075	[.497, .792]
$\sigma_{\theta_5}^2$.590	1.356	.169	[1.012, 1.671]
$\sigma_{\theta_6}^2$.800	1.580	.197	[1.146, 1.952]
ρ_{θ}	.800	.805	.008	[.790, .820]

Table 2.10: Results for the latent traits and item parameters for the unbalanced data study

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

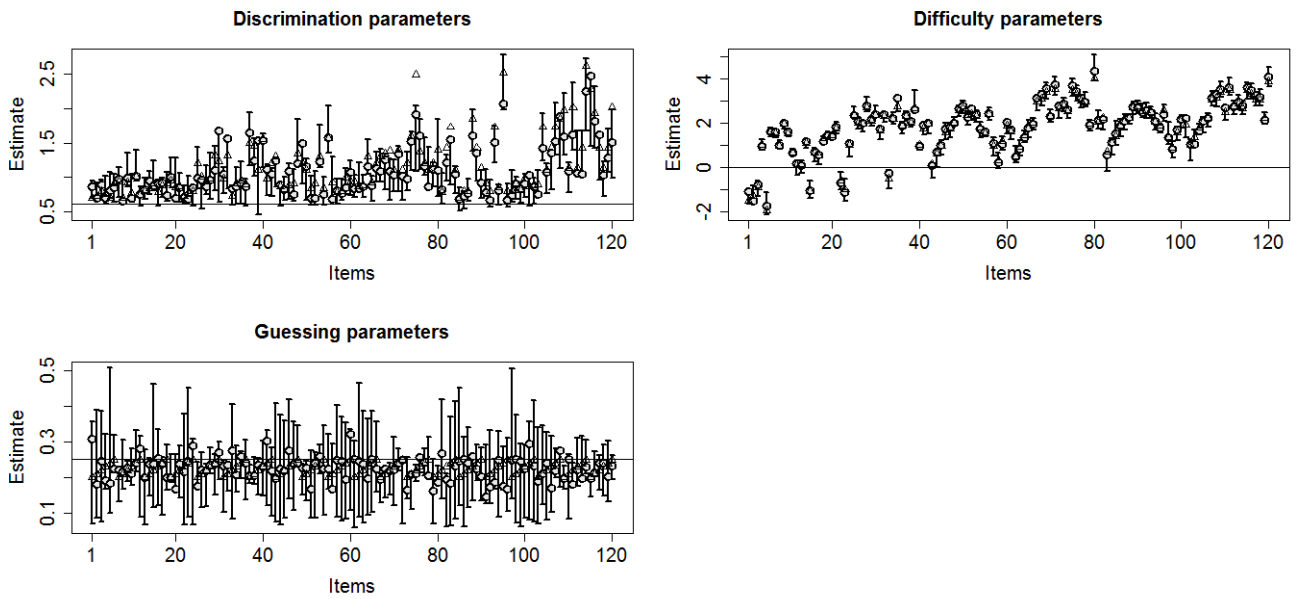


Figure 2.1: Posterior means and 95% credibility intervals for item parameters for ARH(1) model under balanced data. Legend: Circles (estimates), triangles (true values) and Bars (credibility intervals)

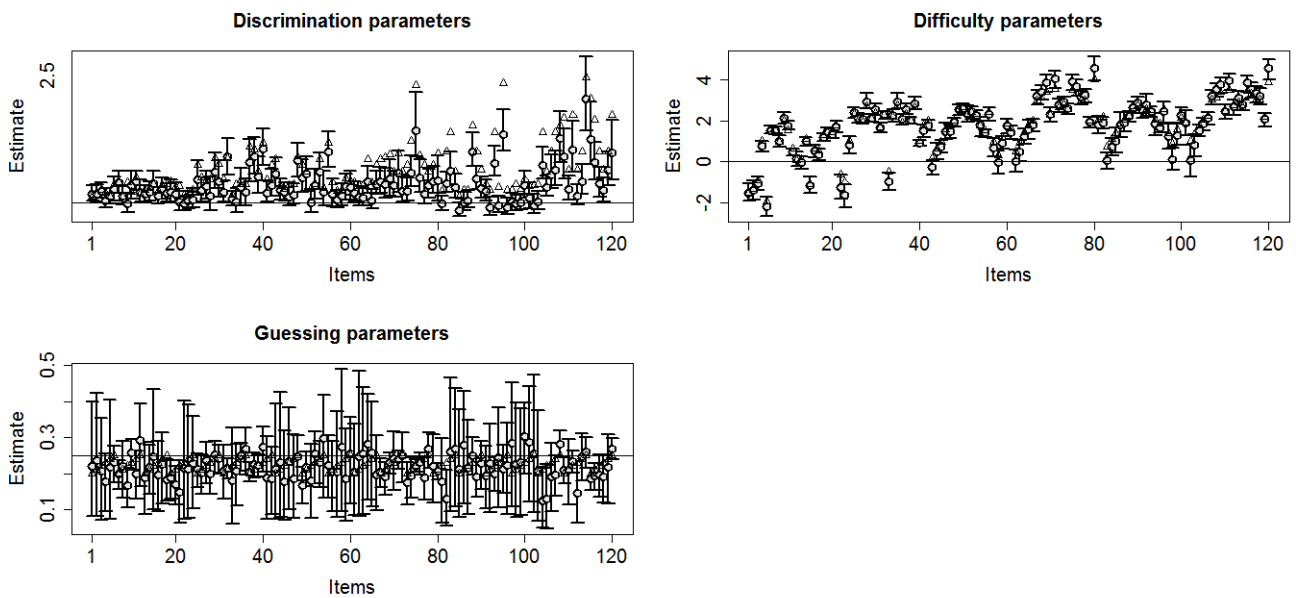


Figure 2.2: Posterior means and 95% credibility intervals for item parameters for ARH(1) model under unbalanced data. Legend: Circles (estimates), triangles (true values) and Bars (credibility intervals)

2.5 Real Data Analysis and Model fit Assessment

2.5.1 Model fit assessment tools

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

$$p(\mathbf{y}_t^{rep}|\mathbf{y}_t^{obs}) = \int p(\mathbf{y}_t^{rep}|\boldsymbol{\vartheta}_t)p(\boldsymbol{\vartheta}_t|\mathbf{y}_t^{obs})d_{\boldsymbol{\vartheta}_t}, \quad (2.5.1)$$

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

$$\mathbb{P}(D(\mathbf{y}_t^{rep}|\boldsymbol{\vartheta}_t) \geq D(\mathbf{y}_t^{obs}|\boldsymbol{\vartheta}_t)|\mathbf{y}_t^{obs}) = \int_{D(\mathbf{y}_t^{rep}) \geq D(\mathbf{y}_t^{obs})} p(\mathbf{y}_t^{rep}|\mathbf{y}_t^{obs})d_{\mathbf{y}_t^{rep}}, \quad (2.5.2)$$

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

For IRT models, Béguin and Glas (2001) have proposed a posterior predictive check to compare the observed score distribution with the posterior predictive score distribution. For the longitudinal IRT model, the observed score distribution can be evaluated per time-point.

To evaluate items fit we defined the following statistic:

$$D_i = \sum_l \frac{|P_{li}^O - P_{li}^E|}{P_{li}^E}, \quad (2.5.3)$$

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

2.5.2 Model Comparison

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

$$D(\vartheta) = -2\text{Log}(L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}_{.}, \boldsymbol{\eta}_{\theta})P(\theta|\boldsymbol{\eta}_{\theta})). \quad (2.5.4)$$

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

$$\overline{D(\vartheta)} = \frac{1}{M} \sum_{m=1}^M D(\vartheta^{(m)}), \quad (2.5.5)$$

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

$$\widehat{\text{DIC}} = D(\bar{\vartheta}) + 2\rho_D, \quad (2.5.6)$$

$$\widehat{\text{EAIC}} = \overline{D(\vartheta)} + 2\rho_D, \quad (2.5.7)$$

$$\widehat{\text{EBIC}} = \overline{D(\vartheta)} + 2\log(n \times I), \quad (2.5.8)$$

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

2.5.3 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 2.11 presents the structure of the tests, that is, the number of items per test and the number of common items across them.

Table 2.11: Structure of tests: real data analyze

	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 longitudinal IRT model considering the unstructured covariance matrix, was applied to the data. The estimated covariance matrix can be seen in Equation (2.5.9). It presents the estimated variances on the main diagonal, estimated correlations on the upper triangular and estimated covariances on the lower triangular. The estimates indicate a time-heteroscedastic structure. Moreover, the correlations are high and decay slowly. Figure 2.3 compares the correlations estimated via unstructured model and the correlations estimated using the structured models considered in the previous section (see Table 2.1). We notice that the correlations estimated by ARH(1) and ARMAH(1,1) decay quickly, and 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.

$$\begin{pmatrix} 1.000 & .848 & .738 & .695 & .677 & .590 \\ .835 & .970 & .816 & .749 & .732 & .649 \\ .626 & .683 & .721 & .818 & .770 & .693 \\ .646 & .686 & .646 & .866 & .846 & .741 \\ .679 & .724 & .657 & .790 & 1.008 & .824 \\ 1.087 & 1.177 & 1.084 & 1.272 & 1.525 & 3.397 \end{pmatrix} \quad (2.5.9)$$

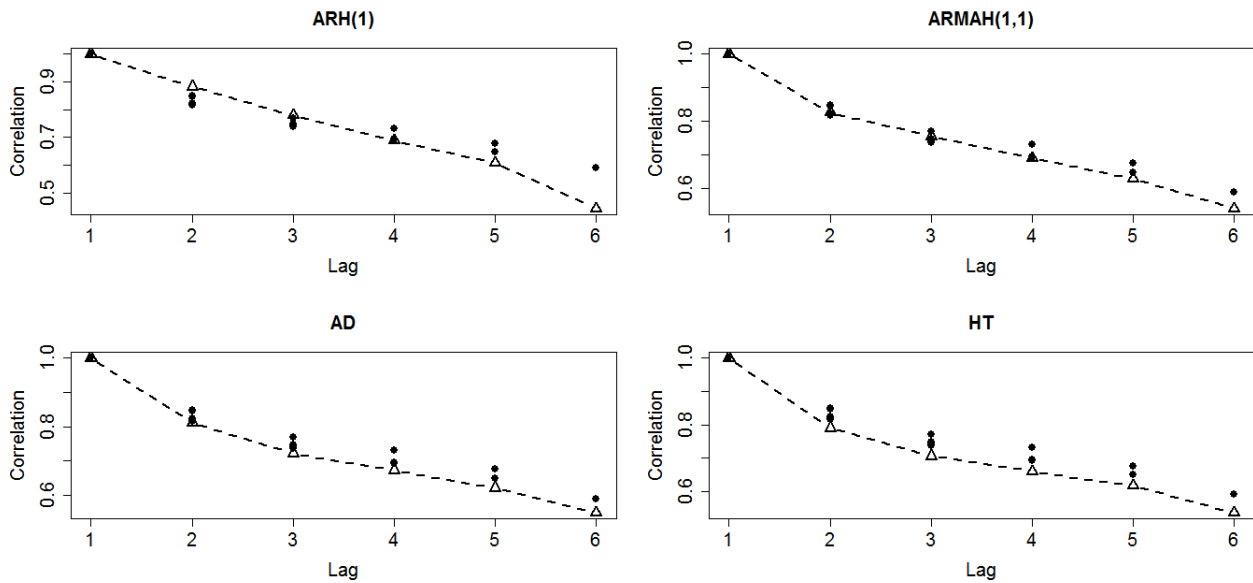


Figure 2.3: Correlation profiles. Legend: unstructured matrix (\bullet), structured matrices ($-\triangle-$)

All the structured covariance models were compared through DIC, EAIC and EBIC. The results are presented in Table 2.12. They indicate that the AD model outperforms the others. Therefore, we will continue the analysis with the selected model.

Table 2.12: Models comparison: Real data analysis

	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

The mean and variance of the first time-point were fixed to zero and one, respectively. Such restriction along with the common items design, ensure the comparability of the latent traits along the tests. Figure 2.4 presents the observed and predicted raw scores with 95% 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 2.5 presents smoothed histograms of the latent trait estimates and the theoretical curves (see Equation (2.2.11)). In general the curves are close, except for grades 4 and 6, which present some difference. Table 2.13 presents the estimates of the population parameters along with the associate posterior standard errors and 95% credibility intervals. The estimates indicate that the mean of the abilities increased from the first to the fifth assessment with a slight decreasing from the fifth to the sixth. It indicates a reasonable improvement on the knowledge 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.

Figure 2.6 presents Bayesian p -values based on the statistic given by Equation (2.5.3) 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 caused by DIF (*Differential*

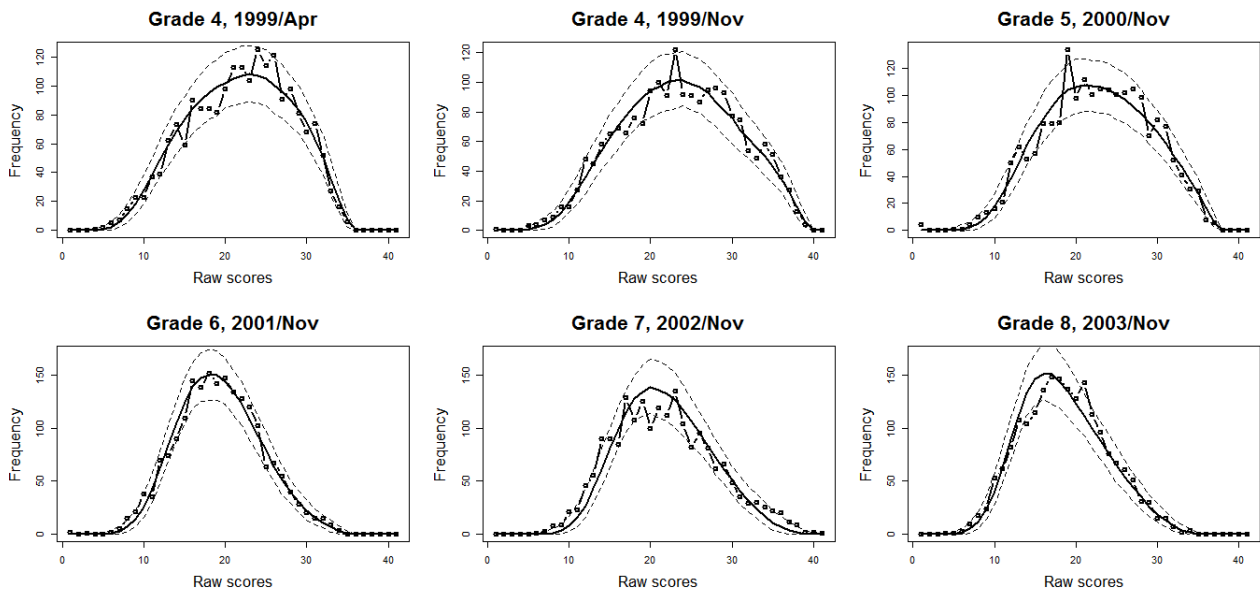


Figure 2.4: Observed score distribution and expected score distribution, and 95% central credibility intervals. Legend: observed scores (dotted line), expected frequency (solid line) and central credibility interval (dashed line)

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.

Figures 2.7 to 2.9 present estimates of the items parameters with their respective 95% credibility intervals. All tests, in general, present a reasonable discrimination power (estimates greater than .6). The discrimination power increase along the tests, except for the last one, where it decreases. It may have been caused by some failure in the study design. Difficulty parameters tend to be higher than the mean of the latent traits along the time-points, indicating that tests are difficult. In general, the guessing parameters estimates indicate that the actual values are different from zero, which supports the use of the three parameters model. Some items presented guessing parameters estimates higher than .3, that is not expected indicating some inconsistency on these items formulation.

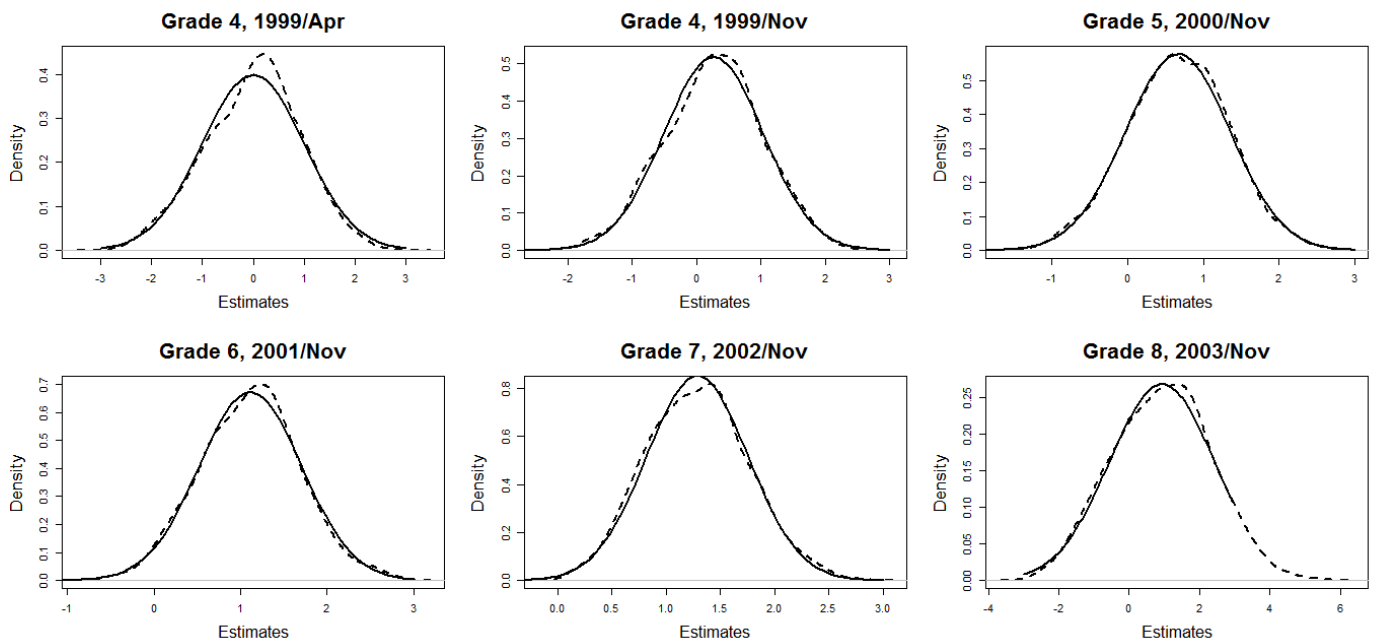


Figure 2.5: Estimated latent traits distributions. Legend: Smoothed histograms (dashed line), Theoretical curve (Solide line)

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 were fitted again and the results can be seen bellow. Figures 2.10 and 2.11 present the predicted scores distribution and the estimated latent trait distributions, respectively. We can see that the grades six, seven and eight, at this time, display a bimodal behavior. This behavior was very likely due the exclusion of the subjects. The items and population parameters estimates were similar to the balanced case.

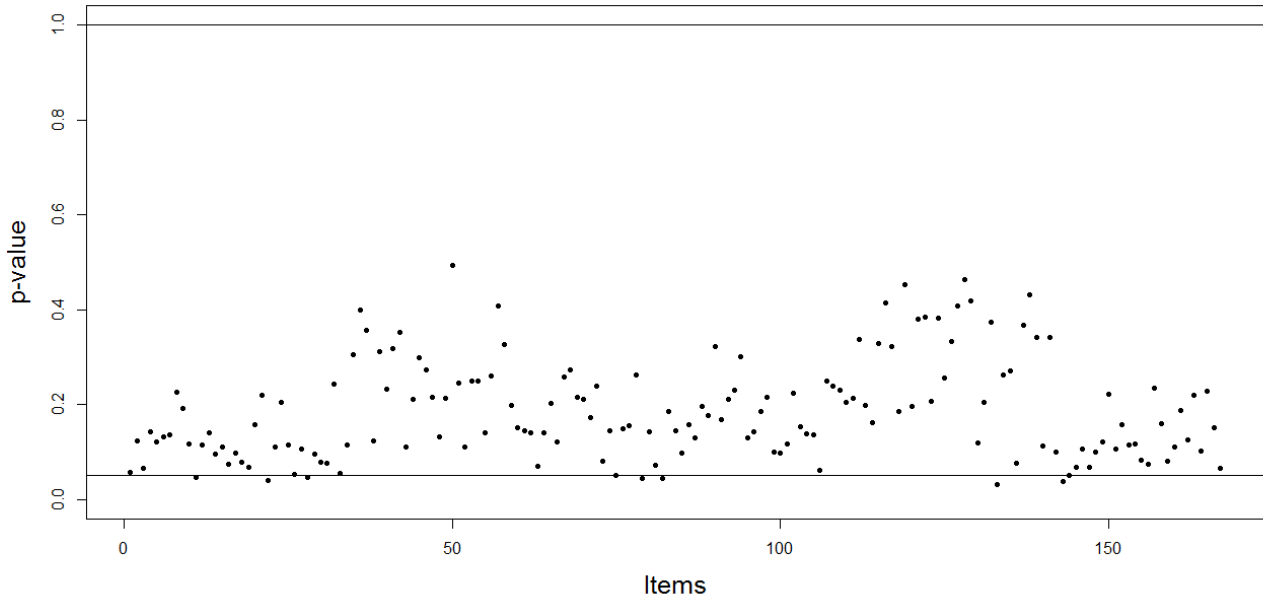
Figure 2.6: Bayesian p -values for item parameters

Table 2.13: 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]				
$\sigma_{\theta_1}^2$	1.000	–	–				
$\sigma_{\theta_2}^2$.597	.057	[.483, .715]				
$\sigma_{\theta_3}^2$.480	.074	[.321, .667]				
$\sigma_{\theta_4}^2$.354	.057	[.242, .472]				
$\sigma_{\theta_5}^2$.219	.032	[.158, .287]				
$\sigma_{\theta_6}^2$	2.255	.610	[1.008, 3.588]				

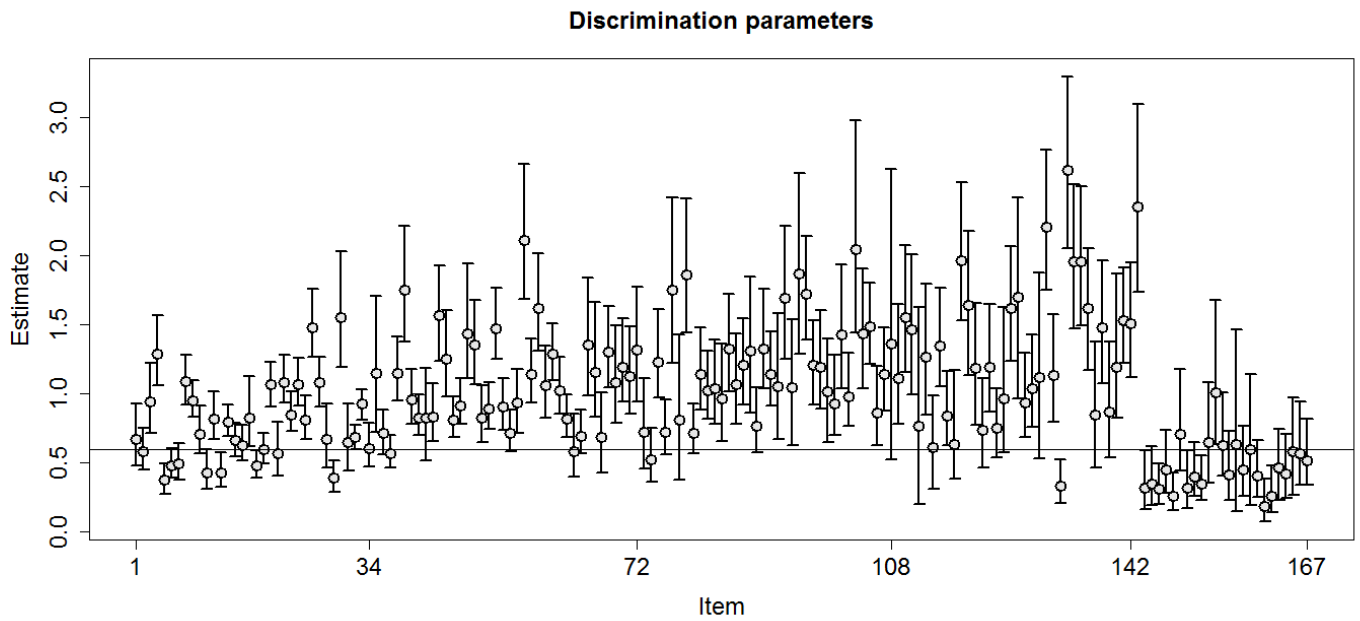


Figure 2.7: Posterior means and 95% central credibility intervals for discrimination parameters

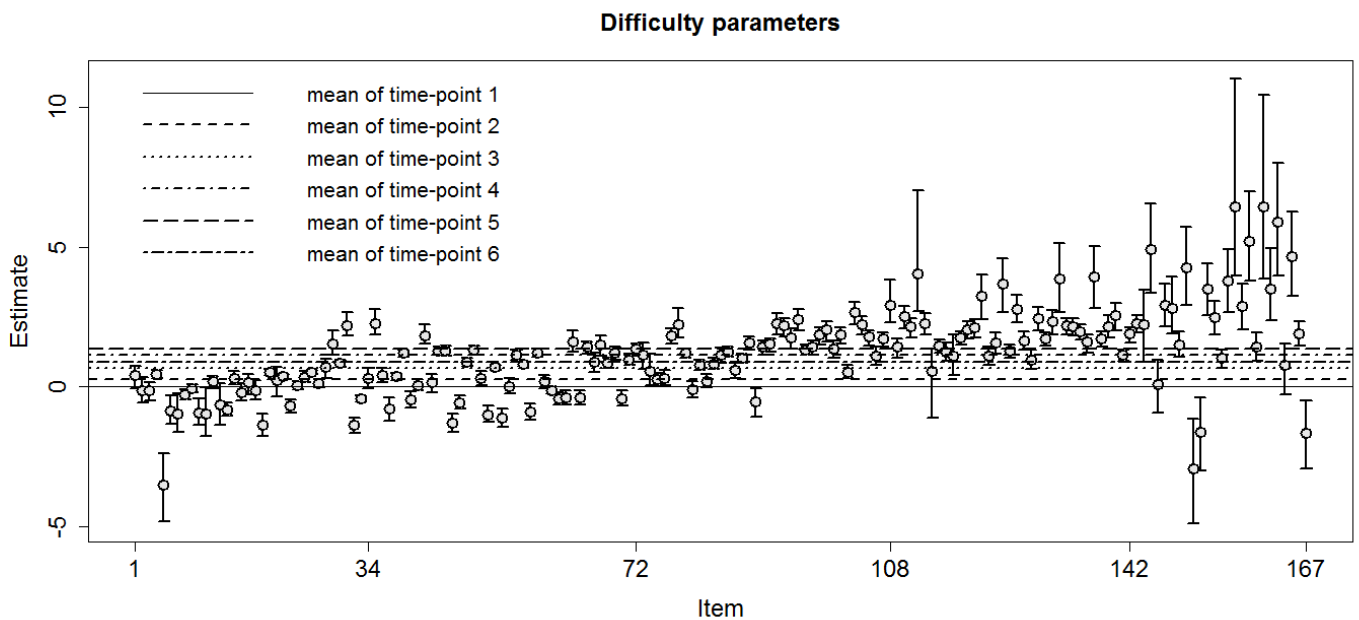


Figure 2.8: Posterior means and 95% central credibility intervals for difficulty parameters

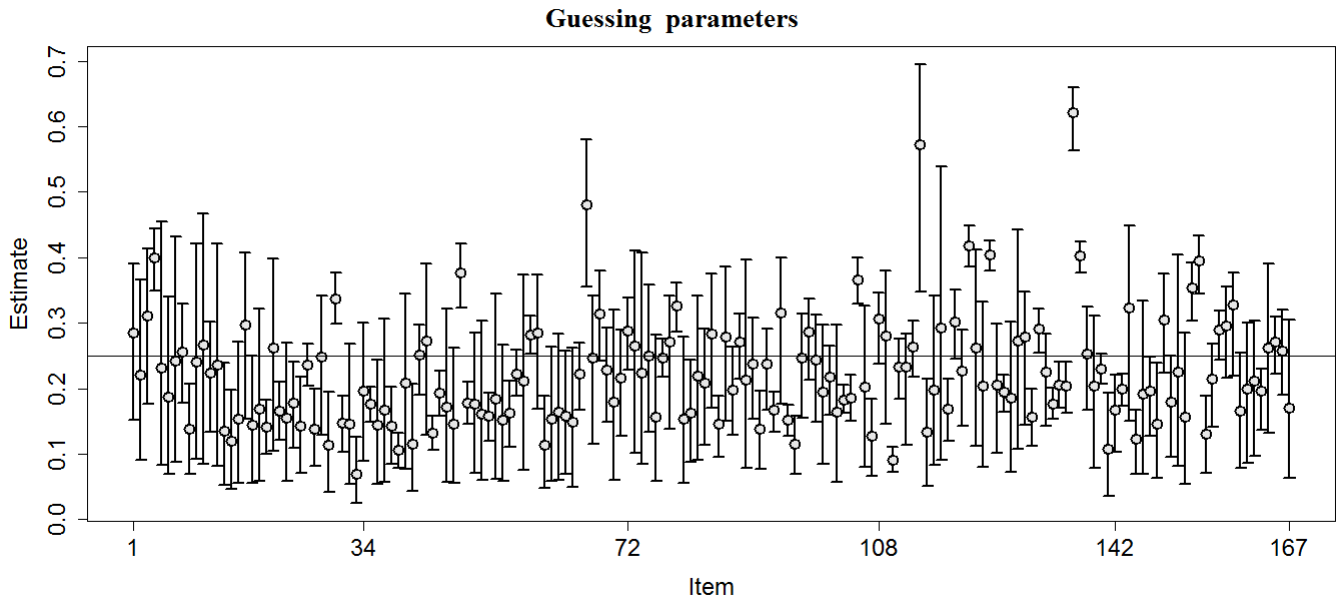


Figure 2.9: Posterior means and 95% central credibility intervals for guessing parameters

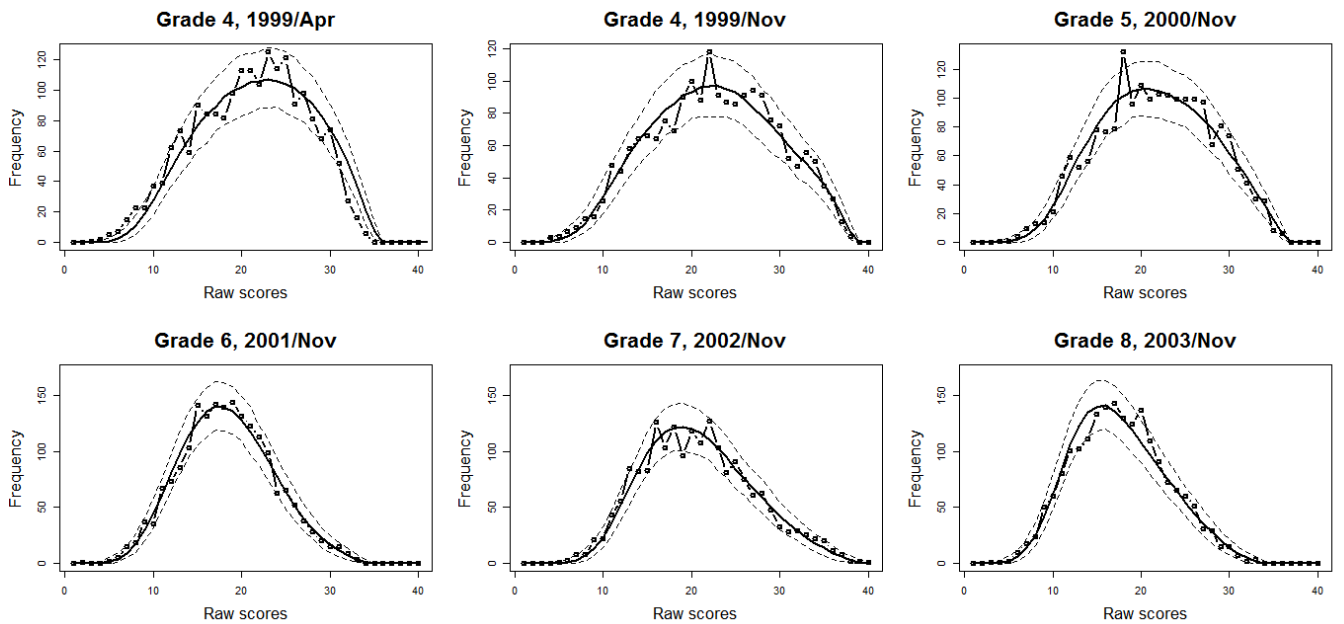


Figure 2.10: Observed score distribution and expected score distribution, and 95% central credibility intervals. Legend: observed scores (dotted line), expected frequency (solid line) and central credibility interval (dashed line). Unbalanced case.

Table 2.14: Estimates of the population parameters according to the AD model (Unbalanced case)

	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]				
$\sigma_{\theta_1}^2$	1.000	–	–				
$\sigma_{\theta_2}^2$.621	.045	[.535, .710]				
$\sigma_{\theta_3}^2$.413	.037	[.343, .486]				
$\sigma_{\theta_4}^2$.325	.031	[.268, .385]				
$\sigma_{\theta_5}^2$.215	.020	[.179, .257]				
$\sigma_{\theta_6}^2$	1.200	.228	[.870, 1.868]				

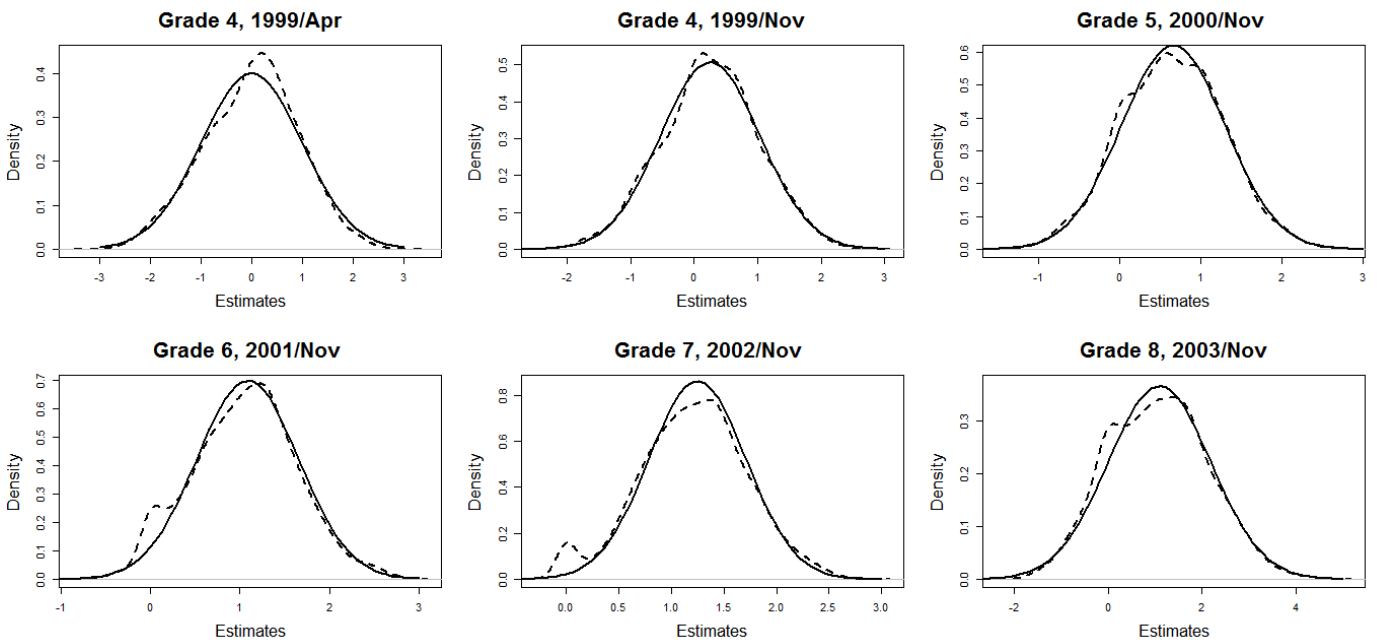


Figure 2.11: Estimated latent traits distributions (Unbalanced case). Legend: Smoothed histograms (dashed line), Theoretical curve (Solide line)

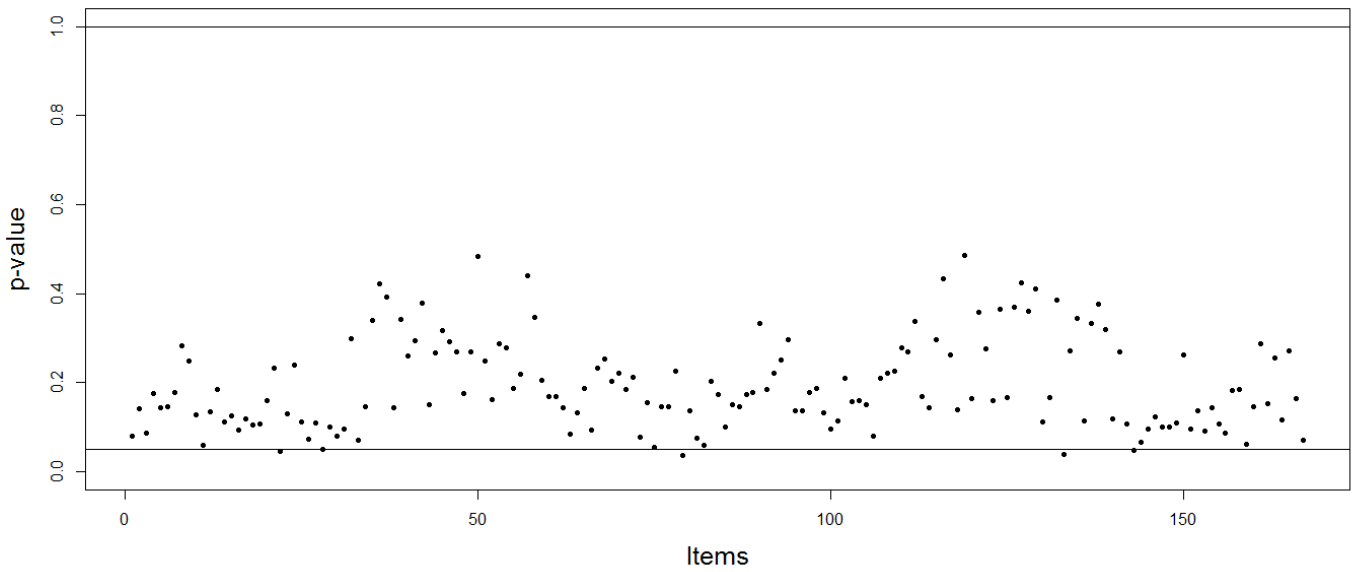


Figure 2.12: Bayesian p -values for item parameters (Unbalanced case).

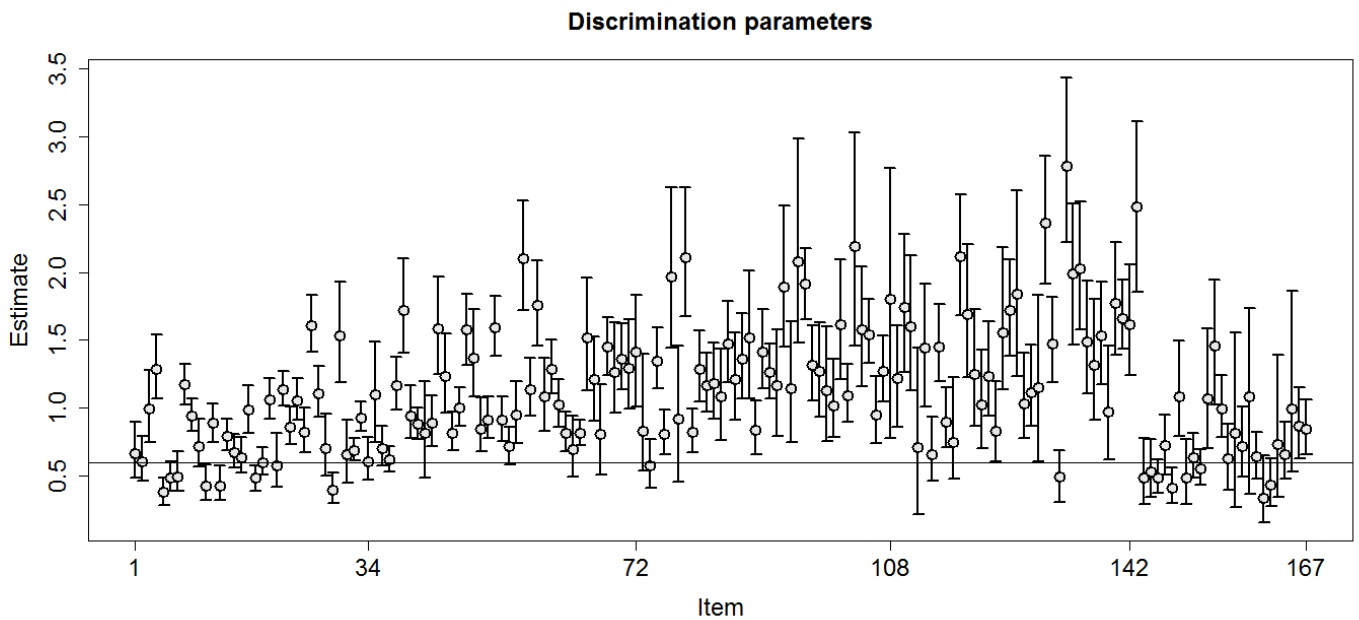


Figure 2.13: Posterior means and 95% central credibility intervals discrimination parameters (Unbalanced case).

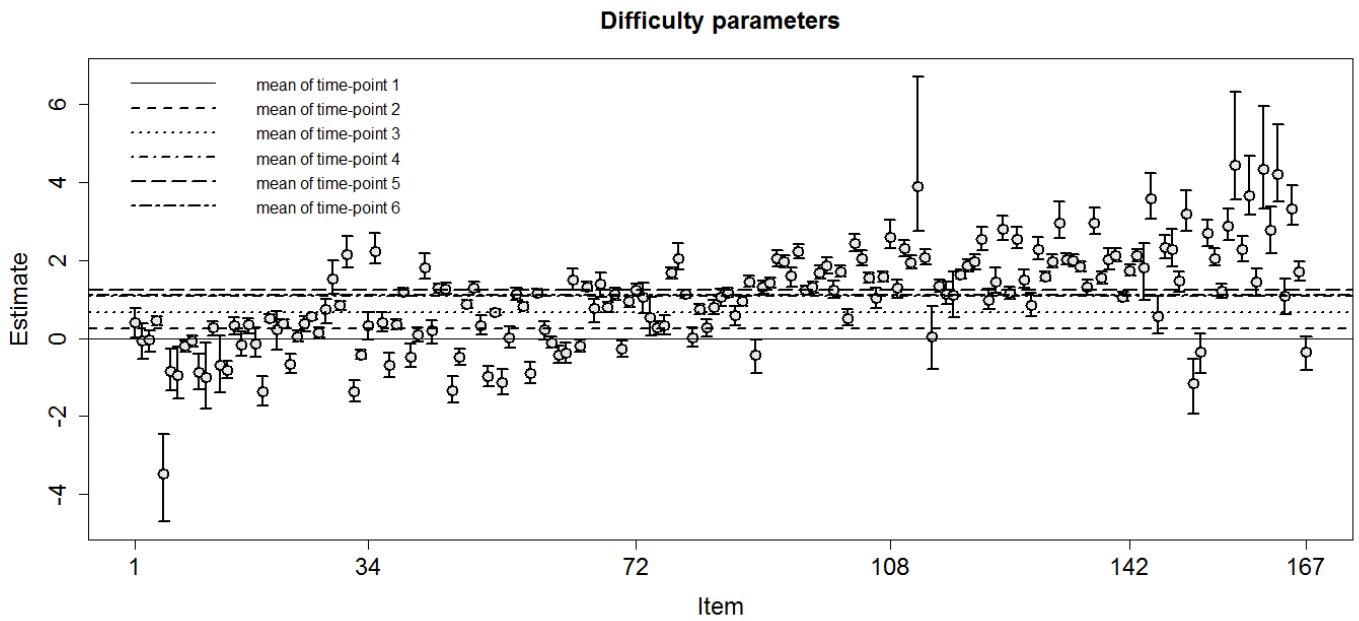


Figure 2.14: Posterior means and 95% central credibility intervals difficulty parameters (Unbalanced case).

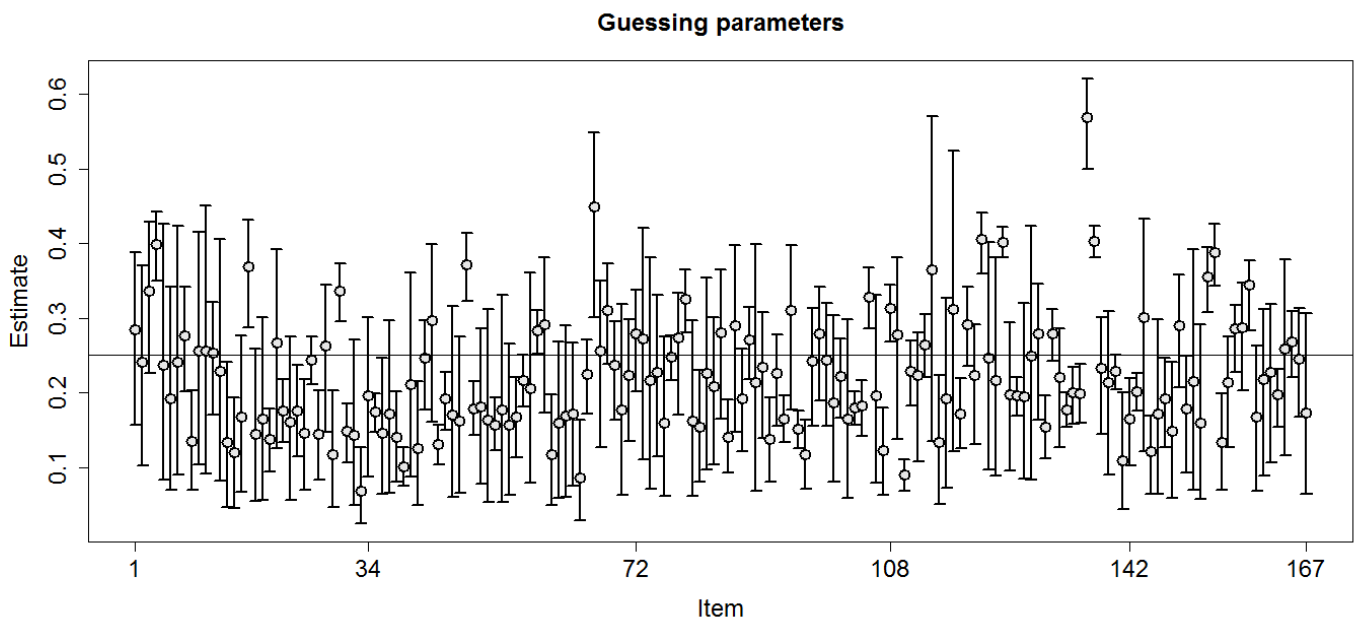


Figure 2.15: Posterior means and 95% central credibility intervals guessing parameters (Unbalanced case).

2.6 Concluding Remarks

We presented a longitudinal three parameters 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, and model fit assessment tools was developed. The FFBS presented a better performance, in terms of autocorrelation reduction, 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 depend 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. In future research we intend to explore some extensions of this model in order to consider other latent trait distributions, growth curves and regression structures for the mean of the latent trait distributions.

Chapter 3

Bayesian general Cholesky decomposition based modeling of longitudinal IRT data with skewed latent distributions.

Abstract

In this chapter we develop a longitudinal IRT model considering skewed latent traits distribution, based on the work of Pourahmadi (1999), which uses the Cholesky decomposition of the matrix of variance and covariance (dependence) of interest related to the latent traits. A kind of multivariate skew-normal distribution for the latent traits is induced by an antedependence model with centered skew-normal errors. We focus on dichotomous responses and considering skewed latent traits 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 (possibly different along these time-points) measuring instrument which have some structure of common items. Using an appropriate augmented data structure, a longitudinal IRT model is developed through the Pourahmadi's approach. The parameter estimation, model fit assessment and model comparison were implemented through a hybrid MCMC algorithm, such that when the full conditionals are not known, the SVE (Single Variable Exchange) 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 methodology developed.

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

3.1 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 of 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 partially or totally different from each other, but they must present some structure of common items. Due to this nested structure, that is, the time-specific measurements within-subjects, it is expected to observe some within-subject dependence. Within 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 will be observed for the latter, see Andrade and Tavares (2005), Azevedo et al. (2012b).

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 (2006a) developed an IRT model to estimate ability distribution parameters, considering several covariance structures, when item parameters are known 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 takes into account important features of the longitudinal data, as heteroscedasticity of population variances and serial correlation.

A common assumption of these longitudinal IRT models is to consider the symmetric normal distribution (either the univariate or the multivariate) to model the latent traits structure. This assumption is often unrealistic and can lead to misleading inference, see Azevedo et al. (2011) and Azevedo et al. (2012a) and references therein. In longitudinal studies, it is very common to observe asymmetry of the marginal latent traits distributions in several time-points, also due to inclusion/exclusion of the subjects along the study. Moreover, from an educational point of view, it is expected to observe a growing and/or decreasing in the latent traits, resulting in negative/positive asymmetry on the latent traits distribution.

Our goal is to develop a general Cholesky decomposition based modeling of longitudinal IRT data under asymmetry of the latent traits distribution. The Antedependence Models, see Pourahmadi (1999) and Nunez-Anton and Zimmerman (2000) were considered. This approach is very flexible and allows to handle multivariate distributions through the univariate conditional distributions. It allows to represent properly a wide range of specific correlation patterns and different latent traits distributions. The centered skew-normal (CSN) distribution (Azzalani, 1985) is considered for the error term, in order to characterize asymmetric behaviors of the latent traits. As pointed out in Azevedo et al. (2012b) the CSN distribution allows to identify the model straightforwardly. In addition, this kind of modeling is quite useful for developing model fit assessment tools.

This chapter is outlined as follows. In Section 3.2 a review of the skew-normal distribution is presented. In Section 3.3, the IRT longitudinal antedependence model is presented along with some of its properties. In Section 3.4, we describe all steps of our MCMC algorithm. In Section 3.5 some simulation studies are conducted to assess the accuracy of our model and MCMC algorithm, concerning some features of interest. In Section 3.6 some model fit assessments are presented and a real data from the Brazilian school development program is analyzed. Finally, in Section 3.7 we presented some conclusions and remarks.

3.2 The Skew-Normal Distribution Under the Centered Parametrization

In order to make this work reasonably self-contained, we begin with an introduction to the skew-normal model. The skew-normal distribution belongs to a subclass of the elliptical distributions (Branco and Arellano-Valle, 2004). It has been used for modeling asymmetric data in many fields, including the psychometrical one, according to Bazán et al. (2004) and Azevedo et al. (2011).

A random variable θ follows the skew-normal (SN) distribution with location parameter $\xi \in \mathbb{R}$, scale parameter $\omega \in \mathbb{R}^+$ and shape parameter $\lambda \in \mathbb{R}$ (notation: $\theta \sim \text{SN}(\xi, \omega, \lambda)$) if its p.d.f is given by

$$p(\theta; \xi, \omega, \lambda) = 2\omega^{-1}\phi\left(\frac{\theta - \xi}{\omega}\right)\Phi\left(\lambda\frac{\theta - \xi}{\omega}\right), \text{ for all } \theta \in \mathbb{R}, \quad (3.2.1)$$

where ϕ and Φ denote the p.d.f and c.d.f of the standard normal distribution, respectively. The mean and variance of θ are given by, respectively

$$\mathbb{E}(\theta) = \xi + \omega\delta r \quad \text{and} \quad \text{Var}(\theta) = \omega^2(1 - r^2\delta^2),$$

where

$$r = \sqrt{\frac{2}{\pi}} \quad \text{and} \quad \delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}.$$

The parameter δ lies in the interval $(-1, 1)$ and it can be used in an alternative parametrization of the SN. Azzalani (1985) introduced the so called centered parametrization defined as follows,

$$\theta_c = \xi + \omega\theta_z = \mu + \sigma\theta_0, \quad (3.2.2)$$

where

$$\theta_z \sim \text{SN}(0, 1, \lambda), \quad \theta_0 = \sigma_z^{-1}(\theta_z - \mu_z)$$

and

$$\mu_z = \mathbb{E}(\theta_z) = r\delta, \quad \sigma_z^2 = \text{Var}(Z) = 1 - \mu_z^2.$$

Then, θ_c follows centered skew-normal distribution (CSN) with parameters defined as:

$$\begin{aligned} \mu = \mathbb{E}(\theta_c) &= \xi + \omega\mu_z, \\ \sigma = \sqrt{\text{Var}(\theta_c)} &= \omega\sqrt{1 - \mu_z^2}, \\ \gamma = \frac{\mathbb{E}[(\theta_c - \mathbb{E}(\theta_c))^3]}{\text{Var}(\theta_c)^{3/2}} &= r\delta^3 \left[\frac{4}{\pi} - 1 \right] [1 - \mu_z^2]^{-3/2}. \end{aligned} \quad (3.2.3)$$

Therefore, the parameters μ and σ^2 are, respectively, mean and variance of the random variable θ_c . The parameter γ stands for the Pearson's asymmetry coefficient and lies in the interval $(-.99527, .99527)$. The closer to -1 or 1 is the parameter γ the negative or positive skewed is the CSN distribution. The CSN distribution is (approximately) symmetric when $\gamma \in (-.13, .13)$. It will be denoted as: $(\theta_c \sim \text{CSN}(\mu, \sigma^2, \gamma))$.

Furthermore, Fisher information matrix obtained through the CSN distribution is nonsingular for all γ and the likelihood is well behaved, unlike the usual skew-normal distribution. For more details,

see Azevedo et al. (2011) and Pewsey (2000).

The p.d.f of the centered skew-normal distribution is given by

$$\begin{aligned} p(\theta_c) &= 2 \frac{\sqrt{\sigma_z^2}}{\sigma} \phi \left[\frac{\sqrt{\sigma_z^2}}{\sigma} \left(\theta_c - \mu + \frac{\sigma}{\sqrt{\sigma_z^2}} \mu_z \right) \right] \Phi \left(\lambda \left[\frac{\sqrt{\sigma_z^2}}{\sigma} \left(\theta_c - \mu + \frac{\sigma}{\sqrt{\sigma_z^2}} \mu_z \right) \right] \right) \\ &= 2\omega^{-1} \phi \left(\omega^{-1}(\theta_c - \xi) \right) \Phi \left[\lambda \omega^{-1}(\theta_c - \xi) \right], \end{aligned} \quad (3.2.4)$$

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

$$\xi = \mu - \sigma \gamma^{1/3} \kappa, \quad (3.2.5)$$

$$\omega = \sigma \sqrt{1 + \gamma^{2/3} \kappa^2},$$

$$\lambda = \frac{\gamma^{1/3} \kappa}{\sqrt{r^2 + \kappa^2 \gamma^{2/3} (r^2 - 1)}}, \quad (3.2.6)$$

where $\kappa = \left(\frac{2}{4-\pi}\right)^{1/3}$. Another important result concerns to the Henze's stochastic representation (Henze, 1986). It means that a random variable $\theta \sim CSN(\mu, \sigma^2, \gamma)$ can be represented by

$$\theta = \xi + \omega \left(\delta X_1 + \sqrt{1 - \delta^2} X_2 \right), \quad (3.2.7)$$

where $X_1 \sim HN(0, 1)$ e $X_2 \sim N(0, 1)$ are independent random variables with half-normal and normal distribution, respectively. Conditional to X_1 we have

$$\begin{aligned} \theta | X_1 &\sim N(\xi + \omega \delta X_1, \omega^2 (1 - \delta^2)), \\ X_1 &\sim HN(0, 1), \end{aligned} \quad (3.2.8)$$

or

$$\begin{aligned} \theta | X_1 &\sim N(\alpha + \tau X_1, \varsigma^2), \\ X_1 &\sim HN(0, 1), \end{aligned} \quad (3.2.9)$$

where $\tau = \omega \delta$ e $\varsigma = \omega \sqrt{1 - \delta^2}$. The last parametrization is more attractive by computational point of view, see Rodriguez (2005).

3.3 Modeling

Following the notation defined in the Chapter 2, 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 properly for dichotomous or dichotomized responses. The second level consists in some kind of multivariate skew-normal distribution induced by the antedependence structure, assuming centered skew-normal erros. That is,

$$\begin{aligned} Y_{ijt} | \theta_{jt}, \zeta_i &\sim \text{Bernoulli}(P_{ijt}), \\ P_{ijt} = \mathbb{P}(Y_{ijt} | \theta_{jt}, \zeta_i) &= c_i + (1 - c_i) \Phi(a_i \theta_{jt} - b_i), \end{aligned} \quad (3.3.1)$$

$$\boldsymbol{\theta}_j | \boldsymbol{\eta}_\theta \sim D_T(\boldsymbol{\eta}_\theta), \quad (3.3.2)$$

where $D_T(\cdot)$ stands for some T -dimensional skew-normal distribution indexed by the parameters $\boldsymbol{\eta}_\theta$. In equation (3.3.1), a_i denote the discrimination parameter, $b_i = a_i b_i^*$, where b_i^* is the original difficulty parameter and c_i is the so called guessing parameter, see Baker and Kim (2004).

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

3.3.1 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 $\text{Cov}(\boldsymbol{\theta}_j) = \boldsymbol{\Sigma}_\theta$.

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

$$\theta_{jt} = \mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}) + \varepsilon_{jt}, \quad t = 1, 2, \dots, T, \quad (3.3.3)$$

where ϕ_{tk} are the so-called generalized autoregressive parameters, see Pourahmadi (1999). The parameters ϕ_{tk} should not be confused with notation ϕ related to the standard normal distribution. In addition, consider $\sum_{k=1}^0 k = 0$.

In matrix form, we have:

$$\boldsymbol{\varepsilon}_j = \mathbf{L}(\boldsymbol{\theta}_j - \boldsymbol{\mu}_\theta) \quad \Leftrightarrow \quad \boldsymbol{\theta}_j = \boldsymbol{\mu}_\theta + \mathbf{L}^{-1}\boldsymbol{\varepsilon}_j. \quad (3.3.4)$$

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

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{21} & 1 & 0 & \cdots & 0 \\ -\phi_{31} & -\phi_{32} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\phi_{T1} & -\phi_{T2} & \cdots & -\phi_{T(T-1)} & 1 \end{pmatrix}. \quad (3.3.5)$$

Then, from (3.3.4) and using the definition of \mathbf{D} we have that,

$$\text{Cov}(\boldsymbol{\varepsilon}_j) = \mathbf{L}\text{Cov}(\boldsymbol{\theta}_j - \boldsymbol{\mu}_\theta)\mathbf{L}' = \mathbf{L}\boldsymbol{\Sigma}_\theta\mathbf{L}' = \mathbf{D}. \quad (3.3.6)$$

Therefore the matrix \mathbf{L} diagonalizes the covariance matrix $\boldsymbol{\Sigma}_\theta$. This result 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 might 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:

$$\Sigma_{\theta} = \begin{pmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta} & \sigma_{\theta_1}\sigma_{\theta_3}\rho_{\theta}^2 \\ \sigma_{\theta_1}\sigma_{\theta_2}\rho_{\theta} & \sigma_{\theta_2}^2 & \sigma_{\theta_2}\sigma_{\theta_3}\rho_{\theta} \\ \sigma_{\theta_1}\sigma_{\theta_3}\rho_{\theta}^2 & \sigma_{\theta_2}\sigma_{\theta_3}\rho_{\theta} & \sigma_{\theta_3}^2 \end{pmatrix}.$$

This is an extension of the *first-order autoregressive* matrix that allow 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 by using equation (3.3.4) we can obtain the following linear model:

$$\begin{aligned} \theta_{j1} - \mu_{\theta_1} &= \varepsilon_{j1}, \\ \theta_{jt} - \mu_{\theta_t} &= \frac{\sigma_{\theta_t}}{\sigma_{\theta_{t-1}}}\rho_{\theta}(\theta_{j(t-1)} - \mu_{\theta_{t-1}}) + \varepsilon_{jt}, \quad t = 2, \dots, T. \end{aligned} \quad (3.3.7)$$

Note that, the parameters $(d_1, d_2, \dots, d_T, \phi_{21}, \phi_{31}, \phi_{32}, \dots, \phi_{T(T-1)})'$ are an one-to-one mapping of the parameters of interest $(\sigma_{\theta_1}^2, \sigma_{\theta_2}^2, \dots, \sigma_{\theta_T}^2, \rho_{\theta})'$. 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 a more complex structured matrix it is very difficult or not possible to obtain a general expression for the antedependence model (as the expression (3.3.7)). However, the Cholesky decomposition can be obtained numerically. Table 3.1 presents some examples of structured matrices.

In order to represent the multivariate structure of the latent traits, we will consider the following antedependence model:

$$\theta_{jt} = \mu_{\theta_t} + \sum_{k=1}^{t-1} \phi_{jk}(\theta_{jk} - \mu_{\theta_k}) + \varepsilon_{jt}, \quad \varepsilon_{jt} \stackrel{i.i.d.}{\sim} CSN(0, d_t, \gamma_{\varepsilon_t}), \quad t = 1, 2, \dots, T. \quad (3.3.8)$$

Then we are considering centered skew-normal distribution for the error term, with mean zero variances d_t and skewness coefficient γ_{ε_t} , that will be called conditional skewness coefficient. The multivariate distribution of the latent trait θ_j of the subject j can be characterized by the density bellow. According to the result (3.2.4) and (3.3.8) we have:

$$\begin{aligned} p(\theta_j | \boldsymbol{\eta}_{\theta}) &= p(\theta_{j1}) \prod_{t=2}^T p(\theta_{jt} | \theta_{j(1:t-1)}) \\ &= 2^T \prod_{t=1}^T \omega_t^{-1} \phi(\omega^{-t}(\theta_{jt} - \beta_{jt})) \Phi[\lambda_t \omega_t^{-1}(\theta_{jt} - \beta_{jt})], \quad j = 1, \dots, n_t, \end{aligned} \quad (3.3.9)$$

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

Table 3.1: Structured covariance matrices used in this work. The σ -parameters are related to variances, while ρ -parameters are related to correlations.

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

$$\begin{aligned} \beta_{j1} &= \xi_1 \\ \beta_{jt} &= \xi_t + \sum_{k=1}^{t-1} \phi_{tk}(\theta_{jk} - \mu_{\theta_k}), \quad j = 2, \dots, n_t. \end{aligned} \quad (3.3.10)$$

To characterize marginal asymmetry, that is, the asymmetry of latent traits at the time-point t , we will also use the Pearson's skewness coefficient. This coefficient is given in terms of the centered moment of the random variable θ_{jt} . The following proposition provides a general expression for the third centered moment.

Proposition 3.3.1. *Consider the antedependence model defined in equation (3.3.8), and let l_{tk} denoting the entries of matrix \mathbf{L}^{-1} . Then the third centered moment of the random variable θ_{jt} is given by:*

$$\mathbb{E}[(\theta_{jt} - \mu_{\theta_t})^3] = d_t^{3/2}\gamma_{\varepsilon_t} + \sum_{k=1}^{t-1} l_{tk}^3 d_k^{3/2}\gamma_{\varepsilon_k}, \quad j = 1, \dots, n_t \text{ and } t = 1, \dots, T. \quad (3.3.11)$$

Proof. By equation (3.3.4) we can write,

$$(\boldsymbol{\theta}_j - \boldsymbol{\mu}_\theta) = \mathbf{L}^{-1} \boldsymbol{\varepsilon}_j.$$

By definition, matrix \mathbf{L} is a lower triangular matrix with ones in the main diagonal (see matrix (3.3.5)). Therefore, \mathbf{L}^{-1} has the same form. Therefore we can write

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

Taking expectations we have,

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

By multinomial theorem, the term $(\varepsilon_{jt} + \sum_{k=1}^{t-1} l_{tk} \varepsilon_{jk})^3$ can be rewrite as:

$$(\varepsilon_{jt} + \sum_{k=1}^{t-1} l_{tk} \varepsilon_{jk})^3 = \sum_{k_1+k_2+\dots+k_t=3} \binom{3}{k_1, k_2, \dots, k_t} \varepsilon_{jt}^{k_t} \prod_{1 \leq m \leq t-1} l_{tm}^{k_m} \varepsilon_{jm}^{k_m}, \quad t = 1, \dots, T,$$

where

$$\binom{3}{k_1, k_2, \dots, k_t} = \frac{3!}{k_1! k_2! \dots k_t!}$$

is a multinomial coefficient. Indices k_1 through k_t are nonnegative integer, such that the sum of all k_m is 3. That is, for each term in the expansion, the exponents of the ε_{jm} must add up to 3. We can also see that,

$$\varepsilon_{jt}^3 + \sum_{k=1}^{t-1} l_{tk}^3 \varepsilon_{jk}^3 + \sum_{\substack{k_1+k_2+\dots+k_t=3, \\ k_m \in \{0,1,2\}}} \binom{3}{k_1, k_2, \dots, k_t} \varepsilon_{jt}^{k_t} \prod_{1 \leq m \leq t-1} l_{tm}^{k_m} \varepsilon_{jm}^{k_m}, \quad t = 1, \dots, T.$$

Now taking expectations,

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

Since ε_{jt} are independent random variables with zero mean, the last term of the sum is equal to zero. Then,

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

Expectations $\mathbb{E}(\varepsilon_{jt}^3)$ are obtained in terms of the conditional skewness parameters. As we defined in equation (3.2.3) the skewness parameter is defined as:

$$\gamma_{\varepsilon_t} = \frac{\mathbb{E}(\varepsilon_{jt}^3)}{d_t^{3/2}} \Leftrightarrow \mathbb{E}(\varepsilon_{jt}^3) = d_t^{3/2} \gamma_{\varepsilon_t}.$$

Thus,

$$\mathbb{E}[(\theta_{jt} - \mu_{\theta_t})^3] = d_t^{3/2} \gamma_{\varepsilon_t} + \sum_{k=1}^{t-1} l_{tk}^3 d_k^{3/2} \gamma_{\varepsilon_k}. \quad (3.3.12)$$

□

The marginal skewness coefficient, denoted by γ_{θ_t} , is defined as: $\gamma_{\theta_t} = \mathbb{E} \left[\left(\frac{\theta_{jt} - \mu_{\theta_t}}{\sigma_{\theta_t}} \right)^3 \right]$. According to proposition 3.3.1 it is given by:

$$\begin{aligned} \gamma_{\theta_1} &= \gamma_{\varepsilon_1} \\ \gamma_{\theta_t} &= \frac{1}{\sigma_{\theta_t}^3} \left[d_t^{3/2} \gamma_{\varepsilon_t} + \sum_{k=1}^{t-1} l_{tk}^3 d_k^{3/2} \gamma_{\varepsilon_k} \right], \quad t = 2, \dots, T, \end{aligned} \quad (3.3.13)$$

where $\sigma_{\theta_t}^2$ are the marginal variances defined as in Chapter 2. That is,

$$\begin{aligned} \sigma_{\theta_1}^2 &= d_1 \\ \sigma_{\theta_t}^2 &= d_t + \sum_{k=1}^{t-1} l_{tk}^2 d_k, \quad t = 2, \dots, T. \end{aligned} \quad (3.3.14)$$

It is important to notice that the multivariate skew-normal (3.3.9) and its respective marginal distributions, are different of those know in the literature, see for example Genton (2004).

3.3.2 Model identification

As in the symmetric case, to ensure identification, we assume that the expectation and standard deviation of the reference time-point (in this case, time-point 1) are fixed, for example, at zero and one, respectively. In other words, we have to ensure that,

$$\theta_{j1} \sim CSN(0, 1, \gamma_{\theta_1}), \quad (3.3.15)$$

with a suitable common items structure along the administered tests. Therefore, the metric (scale) is defined in model (3.3.1) is identified due to the fact that such model is no longer invariant to location-scale transformations, since that the expected value and the standard deviation of the latent distribution of the reference time-point (in this case, time-point 1) are fixed and also due to the linking design. This ensures that the metric for the latent traits is well defined and the results related to all tests (item parameters) and time-points (latent traits and population parameters) lie on the same scale. In addition, the likelihood of our model is much improved compared with the one based on the ordinary skew-normal distribution.

3.4 Bayesian Estimation and MCMC Algorithms

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

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

Consequently, the conditional distribution of W_{ijt} given $Y_{ijt} = y_{ijt}$ corresponds to

$$\begin{aligned} \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 1, \theta_{jt}, \zeta_i) &\propto \Phi(a_i \theta_{jt} - b_i) \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 1, \theta_{jt}, \zeta_i) &\propto c_i (1 - \Phi(a_i \theta_{jt} - b_i)) \\ \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 0, \theta_{jt}, \zeta_i) &= 0 \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 0, \theta_{jt}, \zeta_i) &= 1. \end{aligned} \quad (3.4.1)$$

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

$$Z_{ijt} | (\theta_{jt}, \zeta_i, w_{ijt}) = \begin{cases} N(a_i \theta_{jt} - b_i, 1) \mathbb{I}_{(z_{ijt} \geq 0)}, & \text{if } w_{ijt} = 1, \\ N(a_i \theta_{jt} - b_i, 1) \mathbb{I}_{(z_{ijt} < 0)}, & \text{if } w_{ijt} = 0. \end{cases} \quad (3.4.2)$$

The original data can be represented by

$$Y_{ijt} = \mathbb{I}(Z_{ijt} \geq 0) \mathbb{I}(W_{ijt} = 1) + \mathbb{I}(Z_{ijt} < 0) \mathbb{I}(W_{ijt} = 0), \quad (3.4.3)$$

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

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

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

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

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

$$\begin{aligned} L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta}, | \mathbf{z}_{...}, \mathbf{w}_{...}, \mathbf{y}_{...}) &\propto \prod_{t=1}^T \prod_{j=1}^{n_t} \prod_{i \in I_{jt}} \exp \left\{ -.5(z_{ijt} - a_i \theta_{jt} + b_i)^2 \right\} \mathbb{I}_{(z_{ijt}, w_{ijt})} \\ &\times p(w_{ijt} | y_{ijt}, \theta_{jt}, \zeta_i), \end{aligned} \quad (3.4.4)$$

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 $p(w_{ijt} | y_{ijt}, \theta_{jt}, \zeta_i)$ follows from (3.4.1).

3.4.1 Prior specification and posterior distributions

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

$$p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta} | \boldsymbol{\eta}_{\zeta}, \boldsymbol{\eta}_{\eta}) = \left\{ p(\theta_{j1} | \boldsymbol{\eta}_{\theta_1}) \prod_{t=2}^T \prod_{j=1}^{n_t} p(\theta_{jt} | \theta_{j(1:t-1)}, \boldsymbol{\eta}_{\theta_t}) \right\} \left\{ \prod_{i=1}^I p(\zeta_i | \boldsymbol{\eta}_{\zeta}) \right\} \left\{ \prod_{t=1}^T p(\boldsymbol{\eta}_{\theta_t} | \boldsymbol{\eta}_{\eta}) \right\}, \quad (3.4.5)$$

where the subscript $(1 : t - 1)$ denotes the preceding latent traits. The prior distributions of the latent traits are defined in equation 3.3.8. For the item parameters we have:

$$p(\zeta_{i(-c_i)}) \propto \exp \left[-0.5(\zeta_{i(-c_i)} - \boldsymbol{\mu}_{\zeta})' \boldsymbol{\Psi}_{\zeta}^{-1} (\zeta_{i(-c_i)} - \boldsymbol{\mu}_{\zeta}) \right] \mathbb{I}_{(a_i > 0)} \quad (3.4.6)$$

and

$$c_i \sim \text{Beta}(a_c, b_c), \quad (3.4.7)$$

where $\zeta_{i(-c_i)} = (a_i, b_i)$. In order to obtain full conditional distributions for the population parameters, we consider the Henze's stochastic representation (3.4.8). Therefore, the model (3.3.8) can be rewrite as:

$$\begin{aligned} \theta_{j1} | h_{j1} &\sim N(\xi_1 + \tau_1 h_{j1}, \varsigma_1^2), \\ \theta_{jt} | h_{jt} &\sim N \left(\xi_t + \sum_{k=1}^{t-1} \phi_{tk} (\theta_{jk} - \mu_k) + \tau_t h_{jt}, \varsigma_t^2 \right), \\ H_{jt} &\sim HN(0, 1), \quad t = 1, \dots, T \text{ and } j = 1, \dots, n_t, \end{aligned} \quad (3.4.8)$$

where $\omega_t = \sqrt{\tau_t^2 + \varsigma_t^2}$ and $\lambda_t = \frac{\tau_t}{\varsigma_t}$. The original parameters can be recovered by using equations (3.2.3), that is:

$$\mu_{\theta_t} = \xi_t + r \delta_t \omega_t; \quad d_t = \omega_t^2 (1 - r^2 \delta^2)^2; \quad \gamma_{\varepsilon_t} = r \delta_t^3 \left[\frac{4}{\pi} - 1 \right] \left[1 - r^2 \delta^2 \right]^{-3/2}, \quad (3.4.9)$$

where $\delta_t = \frac{\lambda_t}{\sqrt{1 + \lambda^2}}$. For the Henze's parameters we are considering the following conjugate priors:

$$\begin{aligned} \xi_t &\sim N(\mu_{\xi}, \sigma_{\xi}^2), \\ \tau_t &\sim N(\mu_{\tau}, \sigma_{\tau}^2), \\ \varsigma_t^2 &\sim IG(a_{\varsigma}, b_{\varsigma}), \quad t = 1, \dots, T. \end{aligned} \quad (3.4.10)$$

This parameters will be considered just in the MCMC sampling process and the original parameters (3.4.9) will be recovered later. For the generalized autoregressive parameters we define:

$$\phi_{tk} \sim N(\mu_{\phi}, \sigma_{\phi}^2) \quad t = 2, \dots, T \text{ and } k = 1, \dots, t - 1. \quad (3.4.11)$$

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

$$\rho_{\theta_t} \sim N(\mu_{\rho}, \sigma_{\rho}^2) \mathbb{I}_{[0,1]}, \quad t = 1, 2, \dots, T - 1. \quad (3.4.12)$$

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 (3.4.4) and the prior distribution in equations (3.4.8), (3.4.6), (3.4.7), (3.4.10) and (3.4.11), the joint posterior distribution is given by:

$$\begin{aligned}
p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta}, | \mathbf{z}_{..}, \mathbf{w}_{..}, \mathbf{y}_{..}) &\propto \left\{ \prod_{t=1}^T \prod_{j=1}^n \prod_{i \in I_{jt}} \exp \left\{ -0.5(z_{ijt} - a_i \theta_{jt} + b_i)^2 \right\} p(w_{ijt} | y_{ijt}, \theta_{jt}, \zeta_i) \mathbb{I}(z_{ijt}, w_{ijt}) \right\} \\
&\times \left\{ \prod_{j=1}^{n_t} p(\theta_{j1}, h_{jt}, \boldsymbol{\eta}_{\theta_t}) p(h_{j1}) \prod_{t=2}^T p(\theta_{jt} | \theta_{j(1:t-1)}, h_{jt}, \boldsymbol{\eta}_{\theta_t}) p(h_{jt}) \right\} \\
&\times \left\{ \prod_{i=1}^I p(\zeta_i | \boldsymbol{\eta}_{\zeta}) \right\} \left\{ \prod_{t=1}^T p(\boldsymbol{\eta}_{\theta_t} | \boldsymbol{\eta}_{\eta}) \right\}, \tag{3.4.13}
\end{aligned}$$

where $\boldsymbol{\eta}_{\zeta}$ and $\boldsymbol{\eta}_{\eta}$ are hyperparameters associated with $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}_{\theta}$, respectively, and subscript $(1 : t - 1)$ denotes the preceding latent traits. In addition, we are assuming independence between items and population parameters. Since the posterior distribution has an intractable analytical form, we will use MCMC algorithms in order to obtain empirical approximation for the marginal posterior distribution of interest. In this sense, we can see that the full conditional distribution of the model parameters $\boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta}$ are easy to sample from, at least for the unstructured covariance matrix. For the items parameters we have that:

$$\zeta_{i(-c_i)} | (\cdot) \sim N(\widehat{\Psi}_{\zeta} \widehat{\zeta}_i, \widehat{\Psi}_{\zeta}), \tag{3.4.14}$$

where

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

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

$$c_i | (\cdot) \sim \text{Beta}(s_i + a_c - 1, t_i - s_i + b_c - 1), \tag{3.4.15}$$

where

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

Considering β_{jt} defined as before, for the random effect H_{jt} we have:

$$H_{jt} | (\cdot) \sim N \left[\frac{\tau_t(\theta_{jt} - \beta_{jt})}{\tau_t^2 + \varsigma_t^2}, \frac{\varsigma_t^2}{\tau_t^2 + \varsigma_t^2} \right] \mathbb{I}(h_{jt} > 0). \tag{3.4.16}$$

For the Henze's stochastic representation parameters we have:

$$\xi_t | (\cdot) \sim N(aA; A), \tag{3.4.17}$$

where

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

$$\tau_t | (\cdot) \sim N(bB; B), \quad (3.4.18)$$

where

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

and

$$\zeta_t^2 | (\cdot) \sim IG \left\{ \frac{n_t}{2} + a_\zeta; \frac{1}{2} \sum_{j=1}^{n_t} (\theta_{jt} - \beta_{jt} - \tau_t h_{jt})^2 + b_\zeta \right\}, \quad (3.4.19)$$

where IG stands for the inverse gamma distribution. The generalized autoregressive parameters are simulated by

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

where

$$Q_{tk} = \left(\frac{\sum_{j=1}^{n_t} (\theta_{j(t-1)} - \mu_{\theta_{t-1}})^2}{\zeta_t^2} + \frac{1}{\sigma_\phi^2} \right)^{-1} \quad (3.4.20)$$

$$q_{tk} = \frac{1}{\zeta_t^2} \sum_{j=1}^{n_t} (\theta_{jk} - \mu_{\theta_k}) (\theta_{jt} - \xi_t - \sum_{k \neq t} \phi_{tk} (\theta_{jk} - \mu_{\theta_k}) - \tau_t h_{jt}),$$

for all $t = 2, \dots, T$ and $k = 1, \dots, (t-1)$. In the next sections we will discuss the sampling of the latent traits and correlation parameters of the structured matrices.

3.4.2 The latent traits sampling

A common way to sample from the joint full conditional distribution of θ_j , it is to consider univariate full conditional distributions, in order to sample a time-specific latent trait given all other ones, through Gibbs sampling. It means that: $\theta_{jt} | \theta_{j(-t)}$ for all $t = 1, \dots, T$, where $\theta_{j(-t)}$ is the latent traits vector without the t -th component. That is the idea of the full conditional distribution presented in equation (2.3.11) of the Chapter 2. 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 (Kalman, 1960). It is called **Forward Filtering Backward Sampling** (FFBS).

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:

$$u_t = F_t^\top \theta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \quad (3.4.21)$$

$$\theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, \Omega_t), \quad (3.4.22)$$

where u_t is a sequence of observations at the time, conditionally independent given θ_t and σ_t^2 . 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 model:

$$Z_{jit} = a_i \theta_{jt} - b_i + \xi_{jit}, \quad \xi_{jit} \sim N(0, 1), \quad (3.4.23)$$

$$\theta_{jt} = \beta_{jt} + \tau_t h_{jt} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim N(0, \varsigma_t^2), \quad (3.4.24)$$

where h_{jt} and ε_{jt} are independent and β_{jt} is as defined before with $H_{jt} \sim HN(0, 1)$, for all $i \in I_{jt}$, $j = 1, 2, \dots, n$ and $t = 1, 2, \dots, T$.

FFBS Algorithm

The FFBS algorithm basically consist 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} refer to the information until $t-1$. The system equations (3.4.24) can be written as $\theta_{jt} | \theta_{j(t-1)}, h_{jt} \sim N(\beta_{jt} + \tau_t h_{jt}, \varsigma_t^2)$. By properties of the normal distribution, these specifications can be combined leading to the marginal distribution:

$$\theta_{jt} | \mathbf{z}_j^{t-1}, h_{jt} \sim N(a_{jt}, R_{jt}), \quad (3.4.25)$$

where

$$a_{jt} = \xi_t + \sum_{k=1}^{t-1} \phi_{tk} (m_{jk} - \mu_k) + \tau_t h_{jt} \quad \text{and} \quad R_{jt} = \varsigma_t^2 + \sum_{k=1}^{t-1} \phi_{tk}^2 C_{jk}.$$

Thus,

$$\theta_{jt} | \mathbf{z}_j^t, h_{jt} \sim N(m_{jt}, C_{jt}), \quad (3.4.26)$$

where

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

Equation 3.4.26 is referred in the literature as *Kalman Filter*. Therefore, the backward distributions are given by:

$$\theta_{jt} | \theta_{j(t+1)}, \mathbf{z}_j^t, h_{jt} \sim N(m_{\theta_t}, C_{\theta_t}), \quad (3.4.27)$$

where

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

where $\alpha_{j(t+1)} = \xi_{t+1} - \phi_{t+1,t}\mu_t + \tau_t h_{j(t+1)} + \sum_{k=1}^{t-1} \phi_{t+1,t}(\theta_{jk} - \mu_k)$, for all $j = 1, \dots, n_t$ and $t = 1, \dots, T$. Then, a scheme to sample from the full conditional distribution of θ_j is given by

Algorithm 3.1 FFBS algorithm

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

3.4.3 An alternative correlation parameters sampler

As discussed in chapter 2, an alternative to sample correlation parameters of the structured matrices is to consider the so-called SVE algorithm. This approach is even improved by the oversampling procedure. The same methodology will be used for the model introduced in this Chapter. To illustrate this algorithm consider, for example, an ARH(1) matrix, see Table 3.1. Let $\theta_{..}$ being the set of all latent traits, as defined in Section 3.3 and $p(\theta_{..} | \boldsymbol{\mu}_\theta, \boldsymbol{\phi}, \mathbf{d}, \boldsymbol{\gamma}_\theta)$ denoting the likelihood generated by the antedependence model defined in equation (3.3.8), that is

$$\begin{aligned} p(\theta_{..} | \boldsymbol{\mu}_\theta, \boldsymbol{\phi}, \mathbf{d}, \boldsymbol{\gamma}_\theta) &= \prod_{j=1}^{n_t} p(\theta_{j1} | \mu_{\theta_1}, d_1, \gamma_{\theta_1}) \prod_{t=2}^T p(\theta_{jt} | \mu_{\theta_t}, \boldsymbol{\phi}_t, d_t, \gamma_{\theta_t}) \\ &\propto 2^{Tn_t} \prod_{j=1}^{n_t} \prod_{t=1}^T \omega_t^{-1} \phi_d \left(\omega_t^{-1} (\theta_{jt} - \beta_{jt}) \right) \Phi_d \left[\lambda_t (\omega_t^{-1} (\theta_{jt} - \beta_{jt})) \right], \end{aligned} \quad (3.4.28)$$

where $\boldsymbol{\phi} = (\phi_{21}, \phi_{31}, \phi_{32}, \dots, \phi_{T(T-1)})'$, $\mathbf{d} = (d_1, \dots, d_T)'$, $\boldsymbol{\gamma}_\theta = (\gamma_{\theta_1}, \dots, \gamma_{\theta_T})'$, $\boldsymbol{\mu}_\theta = (\mu_{\theta_1}, \dots, \mu_{\theta_t})$ and $\boldsymbol{\phi}_t$ denotes the elements correspondents to the time-point t . Let us remember that β_{jt} are defined as:

$$\begin{aligned} \beta_{j1} &= \xi_1 \\ \beta_{jt} &= \xi_t + \sum_{k=1}^{t-1} \phi_{jk} (\theta_{jk} - \mu_{\theta_k}), \quad j = 2, \dots, n_t \end{aligned} \quad (3.4.29)$$

and

$$\xi_t = \mu_{\theta_t} - \sqrt{d_t} \gamma_{\theta_t}^{1/3} \kappa, \quad (3.4.30)$$

$$\omega_t = \sqrt{d_t (1 + \gamma_{\theta_t}^{2/3} \kappa^2)}, \quad (3.4.31)$$

$$\lambda_t = \frac{\gamma_{\theta_t}^{1/3} \kappa}{\sqrt{r^2 + \kappa^2 \gamma_{\theta_t}^{2/3} (r^2 - 1)}}, \quad (3.4.32)$$

where r and κ are the constants defined in Section 3.2. Algorithm 3.2 present the correlation parameters sampler for the ARH(1) model for all simulation $m = 1, \dots, M$.

Algorithm 3.2 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 $\boldsymbol{\theta}_{\dots s}^{(m)}$ from the model (3.3.8)

4: Compute $r_1(\boldsymbol{\theta}_{\dots s}^{(m)})$ the first-order sample correlation

5: **end for**

6: Choose the $\rho_{\theta_s}^{(m)}$ and $\boldsymbol{\theta}_{\dots s}^{(m)}$ whose $r_1(\boldsymbol{\theta}_{\dots s}^{(m)})$ is closest to $r_1(\boldsymbol{\theta}_{\dots}^{(m-1)})$

7: Set $\rho_\theta^{(m)} = \rho_{\theta_s}^{(m)}$ and $\boldsymbol{\theta}_{\dots}^{(m)} = \boldsymbol{\theta}_{\dots s}^{(m)}$ the candidate values

8: Build the ARH(1) proposed matrix $\Sigma_{\rho_\theta}^{(m)}$ using `AR1.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**

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

then

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

13: **end if**

For more than one correlation parameter, the algorithm 3.2 can be applied independently to each one by choosing suitable sufficient statistics. It is also possible to sample blocks of correlation parameters. This can be done by modifying line 11 of the algorithm 3.2 to allow accept/reject proposed values jointly. More details about the SVE are found in Appendix A.

In summary, a general algorithm to estimate the parameter's model it is combination of Gibbs sampling with the FFBS and SVE algorithms as we can see in algorithms 3.3 and 3.4.

Algorithm 3.3 Gibbs sampling with FFBS for unstructured matrix

- 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 $Z_{ijt}|\cdot$ for all $i = 1, \dots, I, j = 1, \dots, n$ and $t = 1, \dots, T$.
 - 4: Simulate H_{jt} from $H_{jt}|\cdot$ for all $j = 1, \dots, n$ and $t = 1, \dots, T$.
 - 5: Simulate θ_{jt} using the algorithm 3.1 for all $j = 1, \dots, n$ e $t = 1, \dots, T$.
 - 6: Simulate ζ_i from $\zeta_i|\cdot$ for all $i = 1, \dots, I$.
 - 7: Simulate c_i from $c_i|\cdot$ for all $i = 1, \dots, I$.
 - 8: Simulate ξ_t from $\xi_t|\cdot$ for all $t = 1, \dots, T$.
 - 9: Simulate τ_t from $\tau_t|\cdot$ for all $t = 1, \dots, T$.
 - 10: Simulate ζ_t^2 from $\zeta_t^2|\cdot$ for all $t = 1, \dots, T$.
 - 11: Simulate ϕ_{tk} from $\phi_{tk}|\cdot$ for all $t = 2, \dots, T$ and $k = 1, \dots, t - 1$.
-

Algorithm 3.4 Gibbs sampling with FFBS for structured matrices

- 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 $Z_{ijt}|\cdot$ for all $i = 1, \dots, I, j = 1, \dots, n$ and $t = 1, \dots, T$.
 - 4: Simulate H_{jt} from $H_{jt}|\cdot$ for all $j = 1, \dots, n$ and $t = 1, \dots, T$.
 - 5: Simulate θ_{jt} using the algorithm 3.1 for all $j = 1, \dots, n$ e $t = 1, \dots, T$.
 - 6: Simulate ζ_i from $\zeta_i|\cdot$ for all $i = 1, \dots, I$.
 - 7: Simulate c_i from $c_i|\cdot$ for all $i = 1, \dots, I$.
 - 8: Simulate ξ_t from $\xi_t|\cdot$ for all $t = 1, \dots, T$.
 - 9: Simulate τ_t from $\tau_t|\cdot$ for all $t = 1, \dots, T$.
 - 10: Simulate ζ_t^2 from $\zeta_t^2|\cdot$ for all $t = 1, \dots, T$.
 - 11: Simulate ρ_{θ_i} using a SVE procedure for all necessary correlation parameters.
-

3.5 Parameter recovery study

In this section we study the efficiency of our model and the proposed estimation algorithm concerning parameter recovery. Our algorithm allows to consider different structured covariance matrices. Some examples can be seen in Table 3.1. For simplicity and without lose generality, the AD matrix was chosen to procedure the parameter recovery, since it is the most general matrix considered in this work.

Responses of $n_t = 1500$ subjects, for all t along of $T = 6$ time points were simulated according to the longitudinal model described in Section 3.3, considering the AD matrix. The items 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) (see Table B.1 in appendix). The values of the difficulty parameters were fixed in order to consider low, middle and high difficulty in the items, with respect to the mean of the latent traits along the time-points. Similarly, we fixed high, middle and low discrimination. 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 3.3.8 considering: $\boldsymbol{\mu}_\theta = (.0, 1.0, 1.4, 2.0, 2.3, 2.5)'$, $\boldsymbol{\sigma}_\theta^2 = (1.00, 1.27, .90, .88, .70, .65)'$ and $\boldsymbol{\gamma}_\theta = (.80, .55, .18, .27, .22, -.04)'$ being the vector of marginal skewness coefficients. The correlation parameters was fixed as $\boldsymbol{\rho}_\theta = (.81, .89, .93, .73, .89)'$. 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. (2012b). 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.

Table 3.2 presents the hyperparameters for the prior distributions. Figure 3.1 presents the behavior of the prior distribution of the population parameters. The priors of the population mean, variance and skewness coefficients are presented in form of histogram of their simulated values. The correlation parameter prior correspond to a truncated normal distribution, according to the equation (3.4.12).

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 3.2: Hyperparameters for the prior distributions

Hyperparameters					
μ_ζ	Ψ_ζ	$(\mu_\xi; \sigma_\xi^2)$	$(\mu_\tau; \sigma_\tau^2)$	(a_ζ, b_ζ)	$(\mu_\rho, \sigma_\rho^2)$
(1, 0)	(.5, 16)	(0,10)	(0,10)	(2.1, 1.1)	(0, 10)

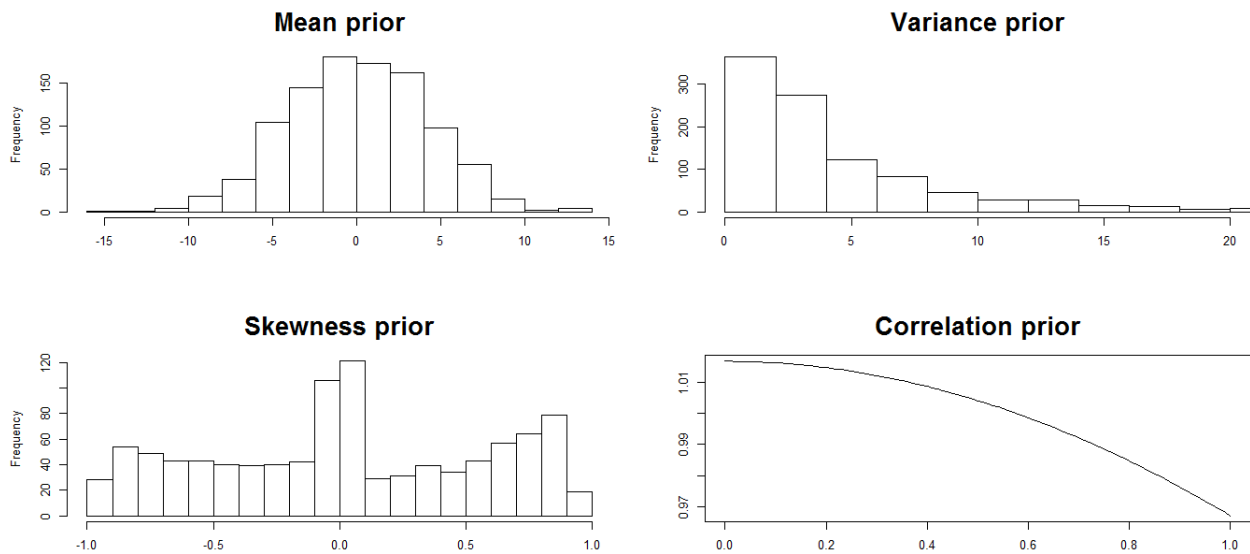


Figure 3.1: Prior distributions of the population parameters

The usual tools to investigate the MCMC algorithms convergence, that is, trace plots, Gelman-Rubin's and Geweke's statistics were monitored. We generate three chains based on three different sets of starting values. The Gelman-Rubin's statistic were close to one for all parameters, indicating convergence. The trace plots and Geweke's monitoring indicate that a Burn-in of 10,000 iterations was enough to reach the convergence. Further, the correlograms indicate that the samples composed by storing every 40th iteration have negligible autocorrelation. Therefore, we will work with valid samples with size 1000.

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

Tables 3.3 and 3.4 present the results of the parameter recovery study. The mean and variance of the first time-point were fixed in 0 and 1, respectively, in order to define the latent trait's scale. This restriction along with the common items design ensure the comparability of the latent traits and the model identification. We can see in Table 3.3 that the estimates of the population parameters are very close to the true values and most of 95% credibility intervals are covering the parameters. Some skewness and correlation parameters present a slight deviation of the true value. Probably this is due to random fluctuations. Table 3.4 presents the results for the latent traits and item parameters. The results indicate that estimates were very accurate. Note that, in the case guessing parameters, the correlation is small. This is expected since the true values have low variability. Figure 3.2 presents the estimates of the latent traits and item parameters with 95% credibility intervals for the item parameters. Considering the item parameters we can see some deviations of the true value only for the discrimination parameter. In a general way, we conclude that the parameters were properly recovered by the estimation algorithm.

Table 3.3: Results for the population parameters.

	True value	Mean	SD	CI(95%)		True value	Mean	SD	CI(95%)
μ_{θ_1}	.000	–	–	–	γ_{θ_1}	.800	.739	.063	[.604, .848]
μ_{θ_2}	1.000	1.011	.039	[.933, 1.088]	γ_{θ_2}	.546	.470	.045	[.373, .547]
μ_{θ_3}	1.400	1.398	.045	[1.315, 1.486]	γ_{θ_3}	.181	.224	.032	[.156, .288]
μ_{θ_4}	2.000	2.057	.069	[1.934, 2.201]	γ_{θ_4}	.275	.244	.027	[.188, .295]
μ_{θ_5}	2.300	2.347	.086	[2.200, 2.543]	γ_{θ_5}	.221	.130	.017	[.098, .167]
μ_{θ_6}	2.500	2.572	.100	[2.402, 2.806]	γ_{θ_6}	-.039	.047	.021	[-.001, .080]
$\sigma_{\theta_1}^2$	1.000	–	–	–	ρ_{θ_1}	.810	.814	.008	[.795, .828]
$\sigma_{\theta_2}^2$	1.270	1.212	.111	[1.022, 1.461]	ρ_{θ_2}	.890	.874	.008	[.859, .886]
$\sigma_{\theta_3}^2$.900	.947	.140	[.750, 1.304]	ρ_{θ_3}	.930	.892	.013	[.864, .916]
$\sigma_{\theta_4}^2$.880	1.050	.188	[.794, 1.536]	ρ_{θ_4}	.730	.801	.021	[.761, .840]
$\sigma_{\theta_5}^2$.700	.984	.193	[.716, 1.470]	ρ_{θ_5}	.890	.838	.027	[.776, .881]
$\sigma_{\theta_6}^2$.650	.956	.189	[.685, 1.445]					

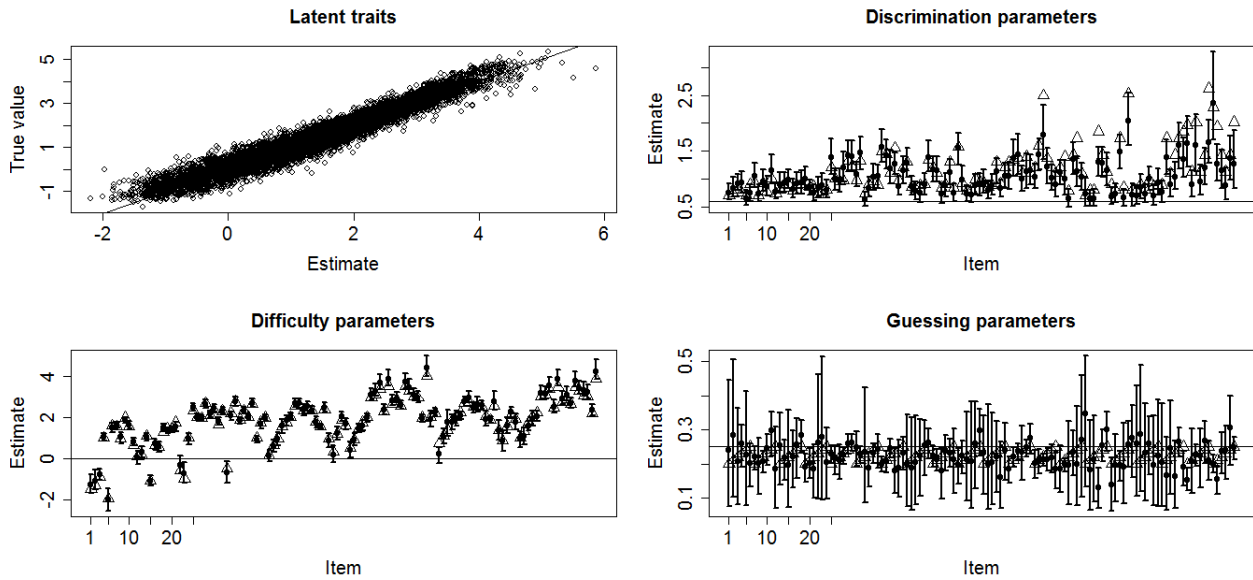


Figure 3.2: Estimates of latent traits and item parameters. Legend: circles denotes the estimates, triangles denotes the true values and the vertical bars denote 95% credibility intervals

Table 3.4: Results for the estimated latent traits and item parameters.

Parameter	Statistic				
	Corr	MBias	MABias	MVAR	MRMSE
Latent trait	.970	-.028	.248	.110	.677
AD Discrimination	.877	.095	.152	.017	.510
Difficulty	.993	-.039	.113	.024	.463
Guessing	.308	.002	.026	.002	.246

3.6 Real Data Analysis and Model fit Assessment

3.6.1 Model fit assessment tools

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

$$p(\mathbf{y}_t^{rep} | \mathbf{y}_t^{obs}) = \int p(\mathbf{y}_t^{rep} | \boldsymbol{\vartheta}_t) p(\boldsymbol{\vartheta}_t | \mathbf{y}_t^{obs}) d_{\boldsymbol{\vartheta}_t}, \quad (3.6.1)$$

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

$$\mathbb{P}(D(\mathbf{y}_t^{rep} | \boldsymbol{\vartheta}_t) \geq D(\mathbf{y}_t^{obs} | \boldsymbol{\vartheta}_t) | \mathbf{y}_t^{obs}) = \int_{D(\mathbf{y}_t^{rep}) \geq D(\mathbf{y}_t^{obs})} p(\mathbf{y}_t^{rep} | \mathbf{y}_t^{obs}) d_{\mathbf{y}_t^{rep}}, \quad (3.6.2)$$

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

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

$$D_i = \sum_l \frac{|P_{li}^O - P_{li}^E|}{P_{li}^E}, \quad (3.6.3)$$

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

3.6.2 Model Comparison

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

$$D(\vartheta) = -2\text{Log}(L(\boldsymbol{\theta}., \boldsymbol{\zeta}., \boldsymbol{\eta}_\theta)P(\theta | \boldsymbol{\eta}_\theta)). \quad (3.6.4)$$

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

$$\overline{D(\vartheta)} = \frac{1}{M} \sum_{m=1}^M D(\vartheta^{(m)}), \quad (3.6.5)$$

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

$$\widehat{\text{DIC}} = D(\bar{\vartheta}) + 2\rho_D, \quad (3.6.6)$$

$$\widehat{\text{EAIC}} = \overline{D(\vartheta)} + 2\rho_D, \quad (3.6.7)$$

$$\widehat{\text{EBIC}} = \overline{D(\vartheta)} + 2\log(n \times I), \quad (3.6.8)$$

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

3.6.3 The 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 2.11 presents the structure of the tests, that is, the number of items per test and the number of common items across them.

Table 3.5: Structure of tests: real data analyze

	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 skew antedependence IRT longitudinal model considering the unstructured covariance matrix, was applied to the data. The estimated covariance matrix can be seen in equation 3.6.9. It presents the estimated variances on the main diagonal, estimated correlations on the upper triangular and estimated covariances on the lower triangular. The estimates indicate a time-heteroscedastic structure. Moreover, the correlations are high and decay slowly. Therefore, due their quick decay, the autoregressive matrices ARH(1) and ARMAH(1,1) are not suitable, as we saw in chapter 2. Unlike the AD and HT structures can better describe the correlation pattern displayed by the data, as we can see in Figure 3.3.

$$\begin{pmatrix} 1.00 & .87 & .77 & .72 & .68 & .57 \\ .71 & .68 & .89 & .83 & .78 & .66 \\ .54 & .51 & .48 & .93 & .88 & .74 \\ .65 & .61 & .58 & .81 & .94 & .79 \\ .75 & .71 & .67 & .93 & 1.22 & .85 \\ 1.17 & 1.11 & 1.05 & 1.46 & 1.91 & 4.18 \end{pmatrix} \quad (3.6.9)$$

We compared skew IRT model under the AD and HT dependence structures with the symmetric IRT longitudinal model with AD structure (the selected model in Chapter 2). For short we will refer the skewed models as skewAD and skewHT and the symmetric AD as AD. Table 3.6 presents the associated statistics for model comparison, where all selected the skew AD model as the best model.

Table 3.6: Statistics for model comparison

	DIC	EAIC	EBIC
AD	496717.2	499933.9	540155.5
skewAD	484481.7	484537.7	485237.4
skewHT	492392.9	494778.6	524609.0

Figure 3.4 presents the observed and predicted scores with 95% credibility intervals for the six time-points. We can see that all the observed scores distribution are well within the intervals, indicating that the model is well fitted. Figures 3.5 and 3.6 present smoothed histograms of the latent trait estimates according to the symmetric and skew AD models, with their respective theoretical curves. The plot of the theoretical curves in the skew case was based on latent traits simulated via model 3.3.8. We can see that, the skew model represents better the latent trait distributions, specially for the time-points 1 and 2. Table 3.7 presents a comparison between the symmetric and skew population parameters estimations. The population variances tend to be smaller according to the skew model, indicating that the symmetric model tends to overestimate the population variance in the presence of asymmetry of the latent traits distribution. In general the standard error for the population parameters were smaller under the skew model. The latent trait distributions at the time-points 1, 2, 3 and 4 presented a moderate asymmetry.

Figure 3.7 presents Bayesian p -values based on the statistic given in Equation (3.6.3) for each item. Items with p -value below .05 or above .90 were assumed not well fitted. We can see that, only few items (around eight items) are not fitted properly. The misfit could be caused by DIF (*Differential item functioning*) or misspecification of the item response function.

Figures 3.8 to 3.10 present the estimates of the items parameters considering both skew and symmetric model. In this Figures circles denote estimates of the skew model, triangles denote estimates of the symmetric model and vertical bars represent 95% credibility intervals. In general we can see that all test presented good discrimination power (estimates greater than .6), except the last one, and difficulty parameters higher than the mean value of the latent traits. Also, we can notice that the discrimination parameters values tend to be higher according to the skew model. Moreover, the guessing parameters estimates indicate that the actual values are different from zero, which supports the use of the three parameters model. Some items presented guessing parameters estimates higher than .3, that is not expected indicating some inconsistency on these items formulation.

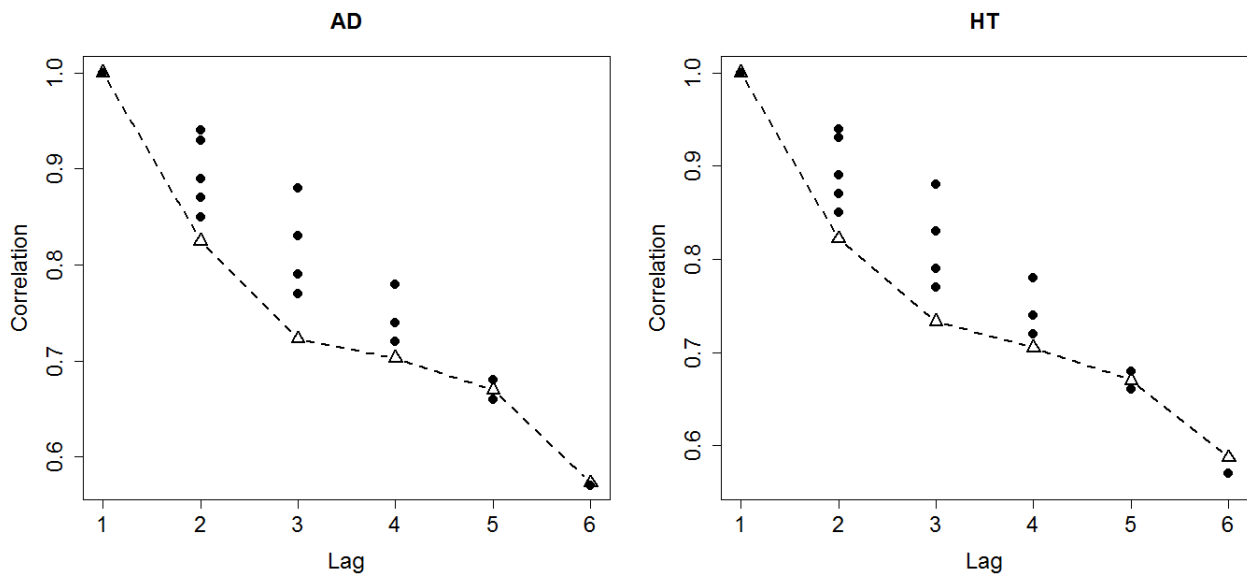


Figure 3.3: Correlation profiles. Legend: unstructured matrix (\bullet), structured matrices ($-\triangle-$).

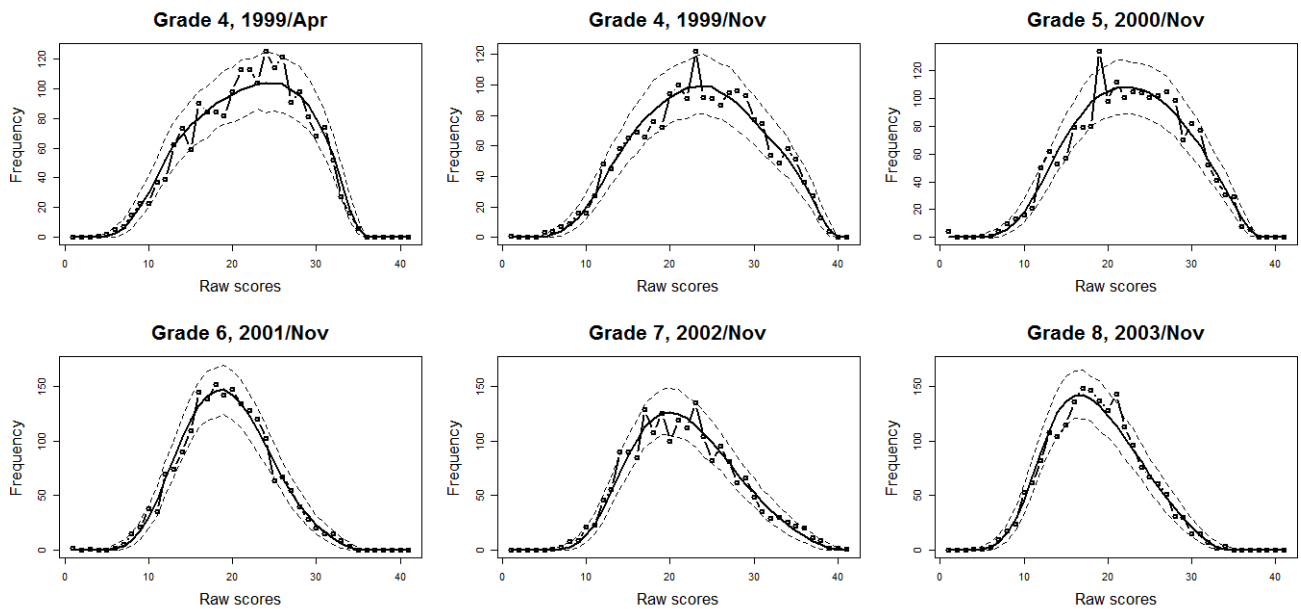


Figure 3.4: Observed and predicted scores distributions with 95% credibility intervals

Table 3.7: Estimates of the population parameters according to the AD model considering both symmetric and skew models

	Symmetric			Skew		
	Mean	SD	CI (95%)	Mean	SD	CI (95%)
μ_{θ_1}	.000	–	–	.000	–	–
μ_{θ_2}	.268	.026	[.222, .320]	.288	.023	[.243, .334]
μ_{θ_3}	.665	.030	[.604, .721]	.644	.023	[.598, .691]
μ_{θ_4}	1.115	.052	[.997, 1.210]	1.029	.036	[.972, 1.103]
μ_{θ_5}	1.290	.065	[1.133, 1.403]	1.160	.046	[1.095, 1.259]
μ_{θ_6}	.935	.074	[.784, 1.073]	1.118	.053	[1.024, 1.218]
$\sigma_{\theta_1}^2$	1.000	–	–	1.000	–	–
$\sigma_{\theta_2}^2$.597	.057	[.483, .715]	.513	.051	[.426, .624]
$\sigma_{\theta_3}^2$.480	.074	[.321, .667]	.391	.061	[.300, .533]
$\sigma_{\theta_4}^2$.354	.057	[.242, .472]	.156	.028	[.111, .221]
$\sigma_{\theta_5}^2$.219	.032	[.158, .287]	.047	.010	[.030, .070]
$\sigma_{\theta_6}^2$	2.255	.610	[1.008, 3.588]	1.429	.213	[1.092, 1.837]
ρ_{θ_1}	.784	.037	[.717, .810]	.824	.005	[.811, .833]
ρ_{θ_2}	.702	.041	[.628, .733]	.740	.008	[.728, .754]
ρ_{θ_3}	.658	.035	[.595, .697]	.687	.012	[.665, .709]
ρ_{θ_4}	.618	.022	[.580, .654]	.627	.011	[.609, .649]
ρ_{θ_5}	.546	.034	[.485, .584]	.567	.013	[.536, .593]
γ_{θ_1}	–	–	–	-.604	.063	[-.724, -.478]
γ_{θ_2}	–	–	–	-.412	.041	[-.489, -.330]
γ_{θ_3}	–	–	–	-.269	.030	[-.330, -.212]
γ_{θ_4}	–	–	–	-.205	.025	[-.252, -.155]
γ_{θ_5}	–	–	–	-.156	.020	[-.195, -.114]
γ_{θ_6}	–	–	–	-.110	.036	[-.185, -.057]

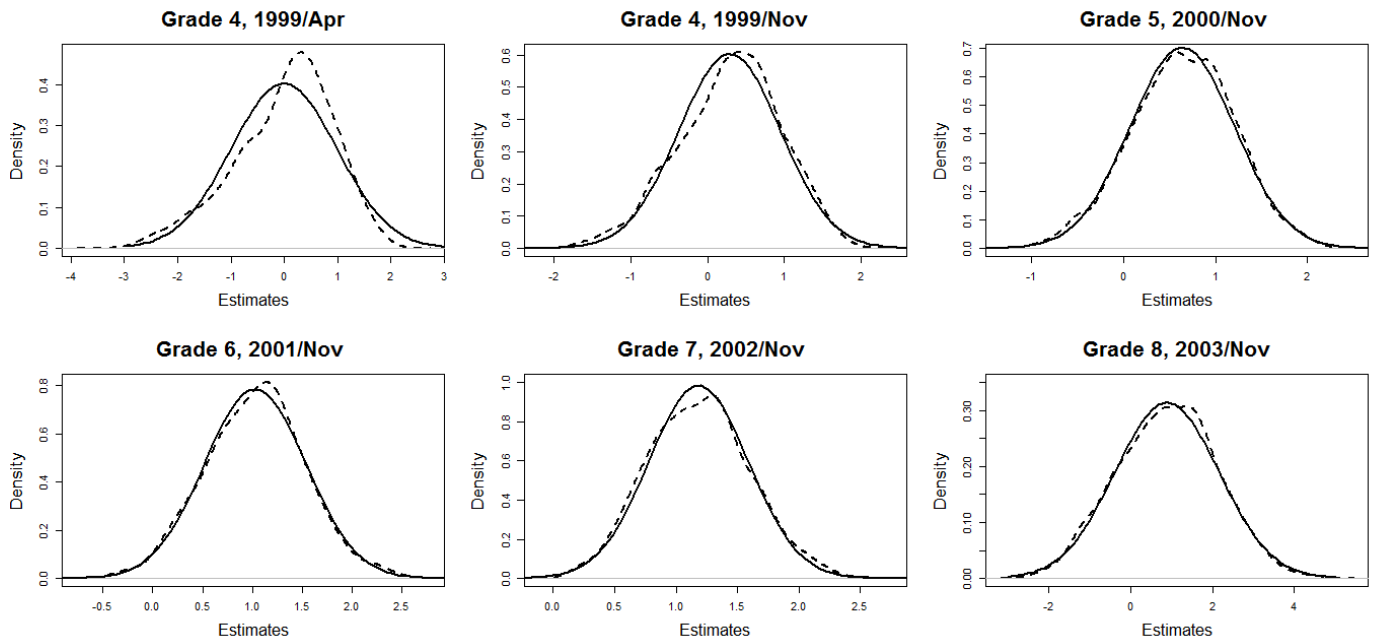


Figure 3.5: Estimated latent traits distributions according to the symmetric model. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)

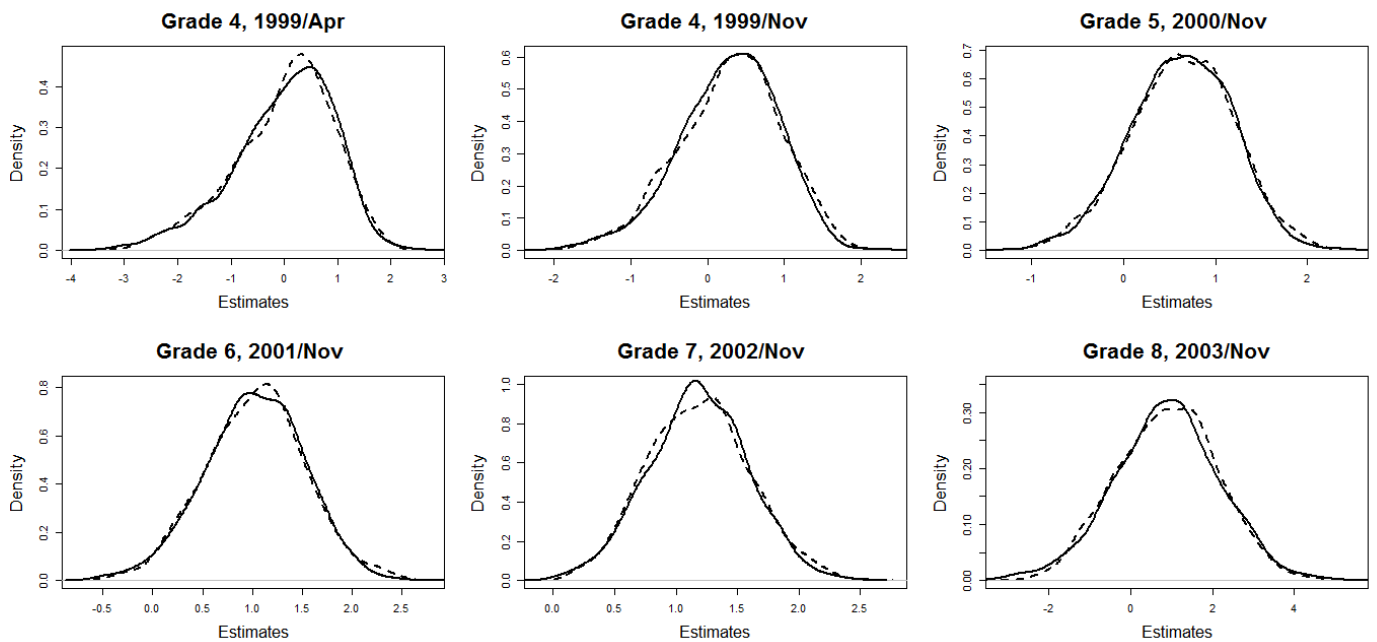


Figure 3.6: Estimated latent traits distributions according to the skew model. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)

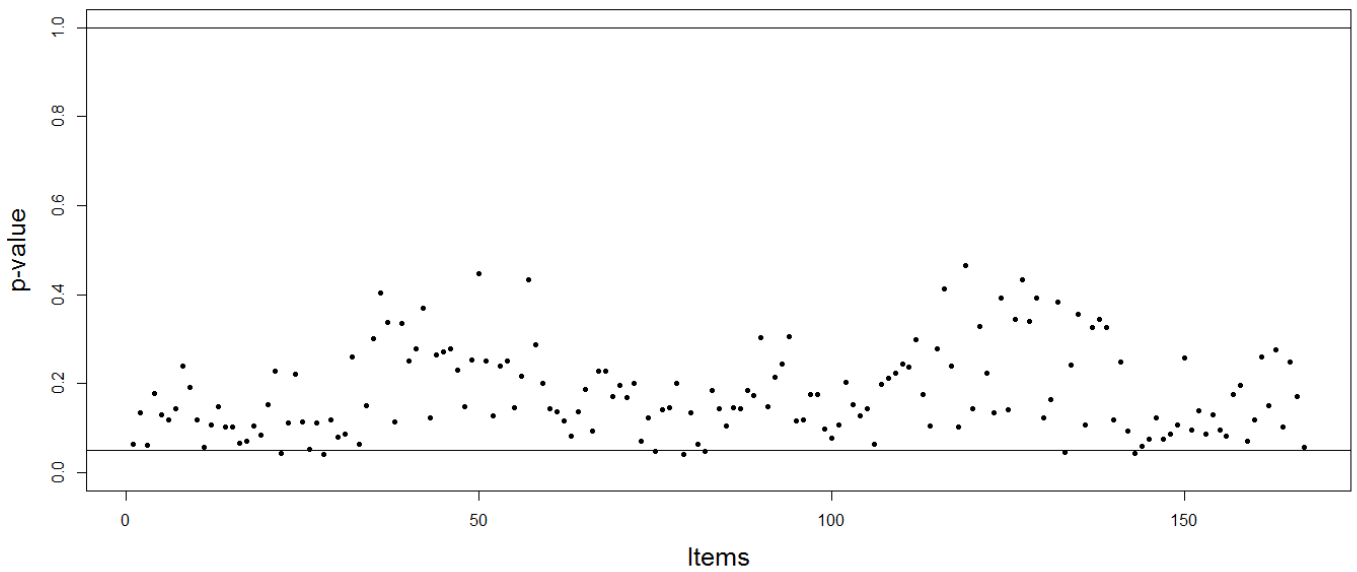


Figure 3.7: Bayesian p -values for the items

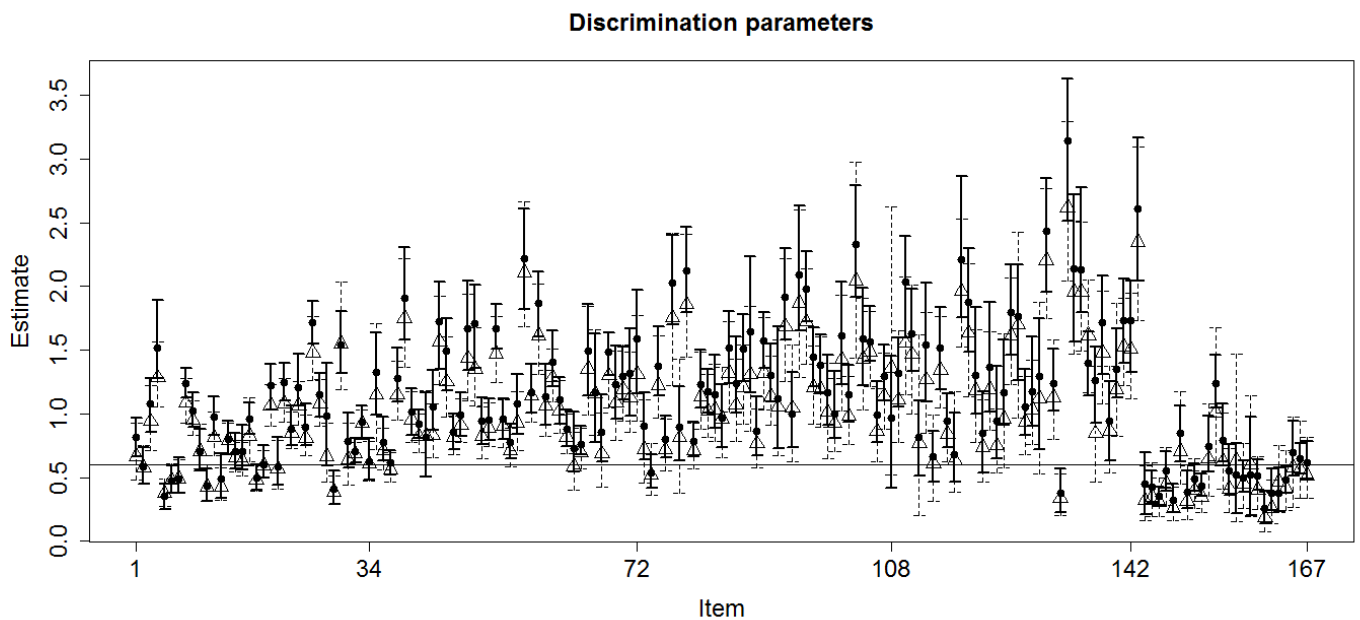


Figure 3.8: Posterior means and 95% central credibility intervals for discrimination parameters.

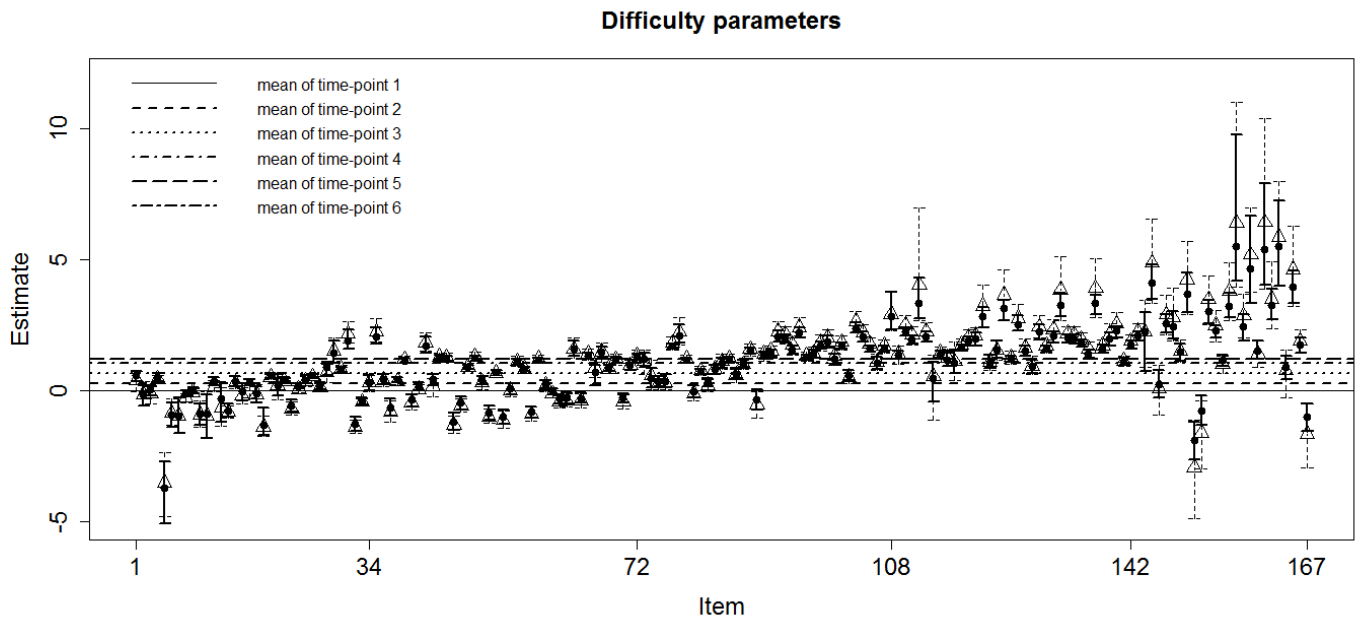


Figure 3.9: Posterior means and 95% central credibility intervals for difficulty parameters.

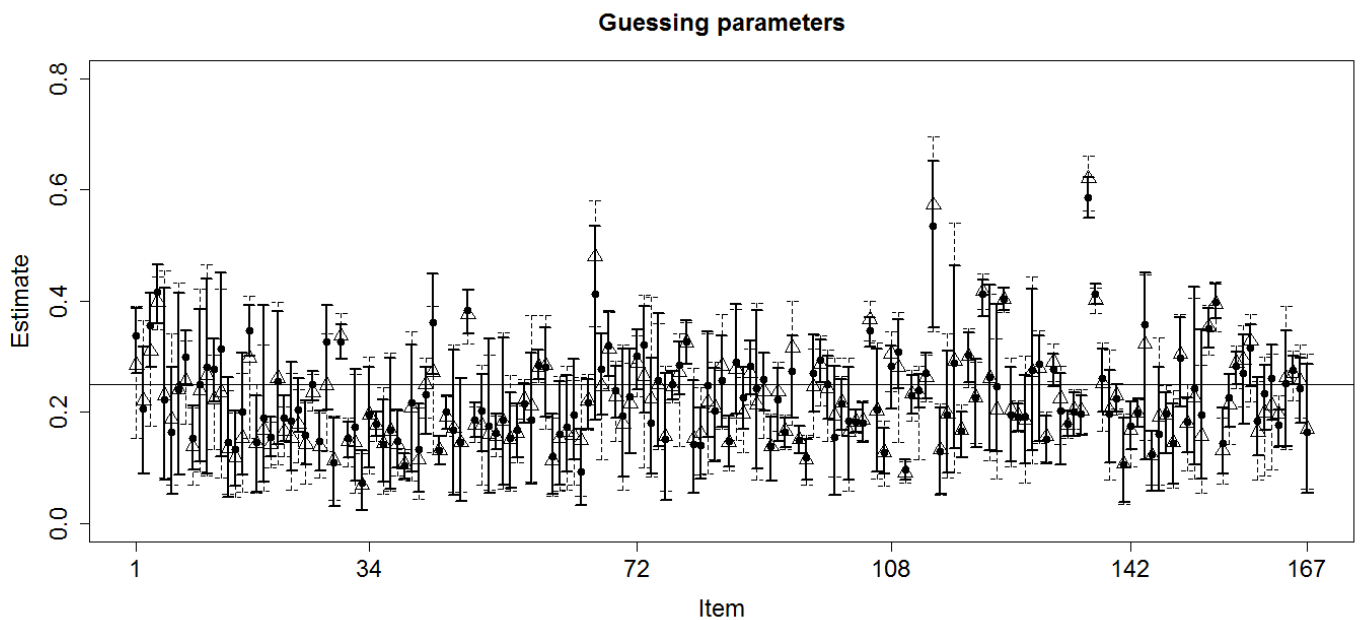


Figure 3.10: Posterior means and 95% central credibility intervals for guessing parameters.

3.7 Concluding Remarks

We presented a longitudinal three parameters IRT model based on a general Cholesky decomposition procedure with skewed latent trait distributions. Such methodology accommodate a wide range of dependence structures and allows asymmetry of the latent distributions. The univariate conditional distributions of the latent traits are assumed to be skew-normally distributed with centered parametrization. We noticed some difficulty to derive the marginal distributions of the latent traits, however, the marginal Pearson's skewness coefficient was relatively easy to obtain. An MCMC algorithm based on the FFBS and SVE procedures was developed for estimating the model parameters. It showed to be efficient in terms of parameter recovery, according to the simulation study. Furthermore, a real data concerning to a Brazilian school development study was analyzed. Some model fit assessment tools were considered, indicating that the model was well fitted. The model identified high between-time correlations of the latent traits. Also, four marginal latent trait distributions presented asymmetric behavior. Further, the skew model fitted better to the data compared to the symmetric model. In conclusion, our approach is a promising alternative to the usual ones in analyzing longitudinal IRT data. In future research we intend to explore some extensions of our model, considering growth curves and regression structures for the population mean of the latent traits distribution. The multiple group structure for longitudinal IRT data could be also considered for the next works as in Azevedo et al. (2015).

Chapter 4

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

Abstract

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

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

4.1 Introduction

Very often in scientific studies involving latent variables there is an interest in studying subjects belonging different groups, along different assessment occasions. For example: the assessment of students from public and private schools followed along scholar grades, the quality of life of men and women measured along weeks, the psychiatric condition of male and female patients along years, among others. This kind of data was named by Azevedo et al. (2015) as Longitudinal Multiple Group Item Response Data. In this kind of IRT data the group heterogeneity can reflect different behaviors, as well as, the longitudinal structure can induce a correlation pattern between the measures of the same subject, which can differ among groups. The multiple group longitudinal IRT model proposed by Azevedo et al. (2015) unifies two methodologies. The multiple group IRT model (Bock and Zimowski, 1997), that allows a simultaneously equating estimation process which leads to more accurate results than a posterior equating, see Kolen and Brennan (2004) and the longitudinal IRT model (Azevedo et al., 2016), that takes into account the within-subject latent trait dependencies. However, this methodology do not consider unbalanced data design caused by dropouts and/or inclusion of subjects, which is very common in longitudinal studies. Another important feature of longitudinal IRT data is the asymmetry of the latent trait distributions induced by inclusion/exclusion of the subjects along the study, growing and/or decreasing in the latent traits, resulting in negative/positive asymmetry, see Azevedo et al. (2011) and Santos et al. (2013) and references therein. The goal of this chapter is to extend the model proposed by Azevedo et al. (2015) in order to handle unbalanced data and asymmetric latent trait distributions. Our model will be build considering a general Cholesky decomposition based modeling of longitudinal IRT data, which are related to the Antependence Models, see Pourahmadi (1999) and Nunez-Anton and Zimmerman (2000). As pointed out in the previous chapter this approach is very flexible and allows to handle multivariate distributions through the univariate conditional distributions. It allows to represent properly a wide range of specific correlation patterns and different latent traits distributions. The centered skew-normal distribution (Azzalani, 1985) is considered for the error term, in order to characterize asymmetric behaviors of the latent trait distributions. In addition, this kind of modeling is quite useful for developing diagnostic tools. We will also consider the modeling of the mean of the latent traits through specific parametric growth curves, which can reduce the number of parameters to be estimated.

This chapter is outlined as follows. In section 4.2 we introduce the model. In section 4.3 we describe the MCMC algorithm developed for parameter estimation. In section 4.4 a parameter recovery study is presented. In section 4.5 some model fit assessment tools are introduced and a real data is analyzed. Finally, in section 4.6 we presented some comments and conclusions.

4.2 Modeling

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

$\mathbf{Y}_{..kt} = (\mathbf{Y}'_{.1t}, \dots, \mathbf{Y}'_{.n_{kt}})'$ is the response vector of all respondents from group k at time-point t and $\mathbf{Y}_{...} = (\mathbf{Y}'_{.1}, \dots, \mathbf{Y}'_{.n})'$ is the entire response matrix. The vector $(y_{ijkt}, \mathbf{y}'_{.jkt}, \mathbf{y}'_{..kt}, \mathbf{y}'_{...})'$ are the respective observed values, respectively. Let $\boldsymbol{\zeta}_i$ be the vector of item parameters of item i , $\boldsymbol{\zeta}$ the vector of all item parameters. The $\boldsymbol{\eta}_{\theta_k}$ is the vector of population parameters of group k and $\boldsymbol{\eta}_{\theta}$ the vector of all population parameters.

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

$$Y_{ijkt} | \theta_{jkt}, \boldsymbol{\zeta}_i \sim \text{Bernoulli}(P_{ijkt}),$$

$$P_{ijkt} = \mathbb{P}(Y_{ijkt} | \theta_{jkt}, \boldsymbol{\zeta}_i) = c_i + (1 - c_i) \Phi(a_i \theta_{jkt} - b_i), \quad (4.2.1)$$

$$\boldsymbol{\theta}_{jk.} | \boldsymbol{\eta}_{\theta} \sim D_{T_k}(\boldsymbol{\eta}_{\theta_k}), \quad (4.2.2)$$

where $D_{T_k}(\cdot)$ stands for some T_k -dimensional skew-normal distribution indexed by the parameters $\boldsymbol{\eta}_{\theta_k}$. In equation (4.2.1), a_i denote the discrimination parameter, $b_i = a_i b_i^*$, where b_i^* is the original difficulty parameter and c_i is the so called guessing parameter, see Baker and Kim (2004).

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

4.2.1 Latent traits modeling

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

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

$$\theta_{jkt} = \mu_{\theta_{kt}} + \sum_{r=1}^{t-1} \phi_{ktr} (\theta_{jkr} - \mu_{\theta_{kr}}) + \varepsilon_{jkt}, \quad t = 1, 2, \dots, T_k \text{ and } k = 1, \dots, K, \quad (4.2.3)$$

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

In matrix form, we have:

$$\boldsymbol{\varepsilon}_{jk.} = \mathbf{L}_k (\boldsymbol{\theta}_{jk.} - \boldsymbol{\mu}_{\theta_k}), \quad (4.2.4)$$

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

matrix $\text{diag}(d_{k1}, d_{k2}, \dots, d_{kT_k})$ and \mathbf{L}_k is a $(T_k \times T_k)$ lower-triangular matrix having the following form,

$$\mathbf{L}_k = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{k21} & 1 & 0 & \cdots & 0 \\ -\phi_{k31} & -\phi_{k32} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\phi_{kT_k1} & -\phi_{kT_k2} & \cdots & -\phi_{T_k(T_k-1)} & 1 \end{pmatrix}. \quad (4.2.5)$$

Then, from (4.2.4) and using the definition of \mathbf{D}_k we have that,

$$\text{Cov}(\boldsymbol{\varepsilon}_{jk.}) = \mathbf{L}_k \text{Cov}(\boldsymbol{\theta}_{jk.} - \boldsymbol{\mu}_{\theta_k}) \mathbf{L}'_k = \mathbf{L}_k \boldsymbol{\Sigma}_{\theta_k} \mathbf{L}'_k = \mathbf{D}_k. \quad (4.2.6)$$

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

More parsimonious models, can be obtained by considering some specific correlation patterns. It can reduce considerably the number of parameters when the restricted correlation pattern is supported by the data, see chapters 2 and 3. The structured matrices considered in this work can be seen in Table 4.1.

Considering centered skew-normal errors for the model (4.2.3) we have,

$$\theta_{jkt} = \mu_{\theta_{kt}} + \sum_{r=1}^{t-1} \phi_{ktr} (\theta_{jkr} - \mu_{\theta_{kr}}) + \varepsilon_{jkt}, \quad \varepsilon_{jkt} \stackrel{i.i.d}{\sim} CSN(0, d_{kt}, \gamma_{\varepsilon_{kt}}), \quad t = 1, 2, \dots, T_k \text{ and } k = 1, \dots, K, \quad (4.2.7)$$

where the CSN distribution is as defined in the previous chapter. The multivariate distribution of the latent traits $\boldsymbol{\theta}_{jk.}$ of the subject j can be characterized by the density bellow. According to the result (3.2.4) of the Chapter 3 and (4.2.7) we have:

$$\begin{aligned} p(\boldsymbol{\theta}_{jk.} | \boldsymbol{\eta}_{\theta_k}) &= p(\theta_{jk1}) \prod_{t=2}^{T_k} p(\theta_{jkt} | \boldsymbol{\theta}_{jk(1:t-1)}) \\ &= 2^{T_k} \prod_{t=1}^{T_k} \omega_{kt}^{-1} \phi \left(\omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt}) \right) \Phi \left[\lambda_t \omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt}) \right], \quad j = 1, \dots, n_{kt}, \end{aligned} \quad (4.2.8)$$

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

$$\beta_{jk1} = \xi_{k1} \text{ and } \beta_{jkt} = \xi_{kt} + \sum_{r=1}^{t-1} \phi_{ktr} (\theta_{jkr} - \mu_{\theta_{kr}}), \quad \text{for all } t = 2, \dots, T_k. \quad (4.2.9)$$

To characterize marginal asymmetry, that is, the asymmetry of latent traits at the time-point t , we will also use the Pearson's skewness coefficient. This coefficient is given in terms of the centered moment of the random variable θ_{jt} . According to proposition 3.3.1 of chapter 3 it is given by:

$$\begin{aligned} \gamma_{\theta_{kt}} &= \gamma_{\varepsilon_{k1}} \\ \gamma_{\theta_{kt}} &= \frac{1}{\sigma_{\theta_{kt}}^3} \left[d_{kt}^{3/2} \gamma_{\varepsilon_{kt}} + \sum_{r=1}^{t-1} l_{ktr}^3 d_{kr}^{3/2} \gamma_{\varepsilon_{kr}} \right], \quad t = 2, \dots, T, \end{aligned} \quad (4.2.10)$$

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

Table 4.1: Structured covariance matrices used in this work. The σ -parameters are related to variances, while ρ -parameters are used for correlations.

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

$$\begin{aligned} \sigma_{\theta_{k1}}^2 &= d_{k1} \\ \sigma_{\theta_{kt}}^2 &= d_{kt} + \sum_{r=1}^{t-1} l_{ktr}^2 d_{kr}, \quad t = 2, \dots, T_k. \end{aligned} \quad (4.2.11)$$

4.2.2 Growth curves for population means

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

$$\boldsymbol{\mu}_{jk.} = (\mu_{\theta_k}(1), \dots, \mu_{\theta_k}(T_k))'; \quad k = 1, \dots, K, \quad (4.2.12)$$

where $\mu_{\theta_k}(\cdot)$ is a parametric function of the time-points. Some curves have been proposed in longitudinal data analysis, see Lindsey (1993) for example. Table 4.2 presents three important curves that

can be useful in many situations. As we discussed in the previous chapters we must to consider some restrictions of the population parameters in order to ensure the identifiability of the model. For the multiple group model the identifiability can be obtained by considering some group as the reference one and then fix the mean and variance of some time-point (reference time-point) of the reference group. In this work we are using a $(0, 1)$ scale, it means that, the mean and variance of the reference group, in the reference time-point is fixed in 0 and 1, respectively. Table 4.3 presents the same curves from Table 4.2 under the restriction $\mu_{\theta_k}(1) = 0$.

Table 4.2: Growth curves for population means

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

Table 4.3: Restricted growth curves for population means

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

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

4.3 Bayesian Estimation and MCMC Algorithms

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

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

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

$$\begin{aligned} \mathbb{P}(W_{ijkt} = 1 | Y_{ijkt} = 1, \theta_{jkt}, \zeta_i) &\propto \Phi(a_i \theta_{jkt} - b_i) \\ \mathbb{P}(W_{ijkt} = 0 | Y_{ijkt} = 1, \theta_{jkt}, \zeta_i) &\propto c_i(1 - \Phi(a_i \theta_{jkt} - b_i)) \\ \mathbb{P}(W_{ijkt} = 1 | Y_{ijkt} = 0, \theta_{jkt}, \zeta_i) &= 0 \\ \mathbb{P}(W_{ijkt} = 0 | Y_{ijkt} = 0, \theta_{jkt}, \zeta_i) &= 1. \end{aligned} \tag{4.3.1}$$

Therefore the augmented variables $\mathbf{Z} = (Z_{1111}, \dots, Z_{1n_{11}11}, \dots, Z_{In_{kT}KT_k})'$, are given by

$$Z_{ijkt} | (\theta_{jt}, \zeta_i, w_{ijkt}) = \begin{cases} N(a_i \theta_{jkt} - b_i, 1) \mathbb{I}_{(z_{ijkt} \geq 0)}, & \text{if } w_{ijkt} = 1, \\ N(a_i \theta_{jkt} - b_i, 1) \mathbb{I}_{(z_{ijkt} < 0)}, & \text{if } w_{ijkt} = 0. \end{cases} \quad (4.3.2)$$

The original data can be represented by

$$Y_{ijkt} = \mathbb{I}(Z_{ijkt} > 0) \mathbb{I}(W_{ijkt} = 1) + \mathbb{I}(Z_{ijkt} \leq 0) \mathbb{I}(W_{ijkt} = 0), \quad (4.3.3)$$

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

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

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

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

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

$$\begin{aligned} L(\boldsymbol{\theta} \dots, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, | \mathbf{z} \dots, \mathbf{w} \dots, \mathbf{y} \dots) &\propto \prod_{k=1}^K \prod_{t=1}^{T_k} \prod_{j=1}^{n_{tk}} \prod_{i \in I_{jkt}} \exp \left\{ -0.5(z_{ijkt} - a_i \theta_{jkt} + b_i)^2 \right\} \mathbb{I}_{(z_{ijkt}, w_{ijkt})} \\ &\times p(w_{ijkt} | y_{ijkt}, \theta_{jkt}, \boldsymbol{\zeta}_i), \end{aligned} \quad (4.3.4)$$

where $\mathbb{I}_{(z_{ijkt}, w_{ijkt})}$ stands for the indicator function $\mathbb{I}_{(z_{ijkt} < 0, w_{ijkt} = 0)} + \mathbb{I}_{(z_{ijkt} \geq 0, w_{ijkt} = 1)}$ and I_{jkt} is the set of items answered by the subject j at time t and $p(w_{ijkt} | y_{ijkt}, \theta_{jkt}, \boldsymbol{\zeta}_i)$ follows from (4.3.1).

4.3.1 Prior and posterior distributions

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

$$p(\boldsymbol{\theta} \dots, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta | \boldsymbol{\eta}_\zeta, \boldsymbol{\eta}_\eta) = \left\{ \prod_{k=1}^K (\theta_{jk1} | \boldsymbol{\eta}_{\theta_{k1}}) \prod_{t=2}^{T_k} \prod_{j=1}^{n_{tk}} p(\theta_{jkt} | \theta_{jk(1:t-1)}, \boldsymbol{\eta}_{\theta_t}) \right\} \left\{ \prod_{i=1}^I p(\boldsymbol{\zeta}_i | \boldsymbol{\eta}_\zeta) \right\} \left\{ \prod_{k=1}^K \prod_{t=1}^{T_k} p(\boldsymbol{\eta}_{\theta_{kt}} | \boldsymbol{\eta}_\eta) \right\}, \quad (4.3.5)$$

where $\boldsymbol{\eta}_\zeta$ and $\boldsymbol{\eta}_\eta$ are hyperparameters associated with $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}_\theta$, respectively, and the subscript $(1 : t - 1)$ denotes the preceding latent traits. In addition, we are assuming independence between

items and population parameters. The prior distributions of the latent traits are defined in equation (4.2.7). For the item parameters we have:

$$p(\zeta_{i(-c_i)}) \propto \exp \left[-.5(\zeta_{i(-c_i)} - \boldsymbol{\mu}_\zeta)' \boldsymbol{\Psi}_\zeta^{-1} (\zeta_{i(-c_i)} - \boldsymbol{\mu}_\zeta) \right] \mathbb{I}_{(a_i > 0)} \quad (4.3.6)$$

and

$$c_i \sim \text{Beta}(a_c, b_c), \quad (4.3.7)$$

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

$$\begin{aligned} \alpha_{kl} &\sim N(\mu_\alpha, \sigma_\alpha^2); \quad l = 1, 2, 3, \\ \sigma_{\theta_{kt}}^2 &\sim IG(a_\sigma, b_\sigma), \\ \gamma_{\varepsilon_{kt}} &\sim N(\mu_\gamma, \sigma_\gamma^2) \mathbb{I}_{[-.99527, .99527]}, \quad k = 1, \dots, K \text{ and } t = 1, \dots, T_k. \end{aligned} \quad (4.3.8)$$

For the generalized autoregressive parameters we define:

$$\phi_{ktr} \sim N(\mu_\phi, \sigma_\phi^2) \quad t = 2, \dots, T_k \text{ and } r = 1, \dots, t - 1. \quad (4.3.9)$$

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

$$\rho_{\theta_{kt}} \sim N(\mu_\rho, \sigma_\rho^2) \mathbb{I}_{[0,1]}, \quad t = 1, 2, \dots, T_k - 1. \quad (4.3.10)$$

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

Given the augmented likelihood in equation (4.3.4) and the prior distribution in equations (4.2.7), (4.3.6), (4.3.7), (4.3.8) and (4.3.9), the joint posterior distribution is given by:

$$\begin{aligned} p(\boldsymbol{\theta}, \dots, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, | \mathbf{z}, \dots, \mathbf{w}, \dots, \mathbf{y}, \dots) &\propto \left\{ \prod_{k=1}^K \prod_{t=1}^{T_k} \prod_{j=1}^{n_{kt}} \prod_{i \in I_{jkt}} \exp \{ -.5(z_{ijkt} - a_i \theta_{jkt} + b_i)^2 \} p(w_{ijkt} | y_{ijkt}, \theta_{jkt}, \boldsymbol{\zeta}_i) \mathbb{I}_{(z_{ijkt}, w_{ijkt})} \right\} \\ &\times \left\{ \prod_{k=1}^K \prod_{j=1}^{n_{kt}} p(\theta_{jk1} | \boldsymbol{\eta}_{\theta_{kt}}) \prod_{t=2}^{T_k} p(\theta_{jkt} | \theta_{jk(1:t-1)}, \boldsymbol{\eta}_{\theta_{kt}}) \right\} \\ &\times \left\{ \prod_{i=1}^I p(\boldsymbol{\zeta}_i | \boldsymbol{\eta}_\zeta) \right\} \left\{ \prod_{k=1}^K \prod_{t=1}^{T_k} p(\boldsymbol{\eta}_{\theta_{kt}} | \boldsymbol{\eta}_\eta) \right\}. \end{aligned} \quad (4.3.11)$$

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

$$\zeta_{i(-c_i)} | (\cdot) \sim N(\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i} \widehat{\boldsymbol{\zeta}}_i, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\zeta}_i}), \quad (4.3.12)$$

where

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

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

$$c_i | (.) \sim \text{Beta}(s_i + a_c - 1, t_i - s_i + b_c - 1), \quad (4.3.13)$$

where

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

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

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

where

$$Q_{tr} = \left(\frac{\sum_{j=1}^{n_{kt}} (\theta_{jk(t-1)} - \mu_{\theta_{k(t-1)}})^2}{\zeta_{kt}^2} + \frac{1}{\sigma_\phi^2} \right)^{-1} \quad (4.3.14)$$

$$q_{tr} = \frac{1}{\zeta_{kt}^2} \sum_{j=1}^{n_{kt}} (\theta_{jkr} - \mu_{\theta_{kr}}) (\theta_{jkt} - \xi_{kt} - \sum_{r \neq t} \phi_{tr} (\theta_{jkr} - \mu_{\theta_{kr}}) - \tau_{kt} h_{jkt}),$$

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

Unlike to the model of the Chapter 3, the Henze's stochastic representation is not useful to sample from population means, variances and skewness coefficients. since the population means are modeled by growth curves, the reparametrization presented in Chapter 3 is no longer possible. As a consequence, even using the Henze's stochastic representation is no longer possible neither to use that reparametrization neither to obtain known full conditional. Therefore some auxiliary algorithm is necessary. The Metropolis-Hastings algorithm is considered using the original likelihood of the skew-normal distribution. Let us denote the conditional distribution of $\theta_{jt} | \theta_{j(1:t-1)}, \boldsymbol{\eta}_{\theta_{kt}}$ as

$$L(\theta_{jkt} | \theta_{j(1:t-1)}, \boldsymbol{\eta}_{\theta_{kt}}) \equiv L(\theta_{jkt} | \boldsymbol{\eta}_{\theta_{kt}}) = 2^{T_k} \omega_{kt}^{-1} \phi \left(\omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt}) \right) \Phi \left[\lambda_{kt} (\omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt})) \right]. \quad (4.3.15)$$

The function β_{jkt} is defined in equations (4.2.9). In order to perform the Metropolis-Hastings steps we need to defined the following proposal densities:

$$\begin{aligned} p(\alpha_{kl}^{(m)}, \alpha_{kl}^{(m-1)}) &= N(\alpha_{kl}^{(m-1)}, \sigma_0^2); \quad l = 1, 2, 3, \\ p((\sigma_{\theta_k}^2)^{(m)}, (\sigma_{\theta_k}^2)^{(m-1)}) &= \text{Lognormal}((\sigma_{\theta_k}^2)^{(m-1)}, \sigma_0^2) \\ p(\gamma_{\varepsilon_{kt}}^{(m)}, \gamma_{\varepsilon_{kt}}^{(m-1)}) &= U \left(\nu_1 (\gamma_{\varepsilon_{kt}}^{(m-1)}), \nu_2 (\gamma_{\varepsilon_{kt}}^{(m-1)}) \right); \quad k = 1, \dots, K, \end{aligned} \quad (4.3.16)$$

where

$$\nu_1 = \max \left\{ -0.99527, \gamma_{\varepsilon_{kt}}^{(m-1)} - \Delta_\gamma \right\} \quad \text{and} \quad \nu_2 = \max \left\{ 0.99527, \gamma_{\varepsilon_{kt}}^{(m-1)} + \Delta_\gamma \right\}, \quad \Delta_\gamma > 0.$$

The constant Δ_γ is defined in a suitable value, see Azevedo et al. (2011). The superscript (m) indicate the estimate at the iteration m of the MCMC algorithm. The algorithms 4.1 to 4.3 describe the metropolis-hastings steps for each growth curve, variance and conditional skewness parameters.

Algorithm 4.1 Growth curve parameters sampler

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

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

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

Algorithm 4.2 Variance parameters sampler

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

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

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

Algorithm 4.3 Conditional skewness parameters sampler

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

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

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

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

The latent trait sampler

To sample latent traits we consider the Henze's stochastic representation of the skew-normal distribution discussed in Chapter 3. Therefore, considering the augmented data structure, we have the following representation in terms of dynamic model:

$$Z_{jikt} = a_i \theta_{jkt} - b_i + \xi_{jikt}, \quad \xi_{jikt} \sim N(0, 1), \quad (4.3.17)$$

$$\theta_{jkt} = \beta_{jkt} + \tau_{kt} h_{jkt} + \varepsilon_{jkt}, \quad \varepsilon_{jkt} \sim N(0, \varsigma_{kt}^2), \quad (4.3.18)$$

where h_{jkt} and ε_{jkt} are independent and β_{jt} is define as before with $H_{jkt} \sim HN(0, 1)$. Then the forward step of the FFBS is described below. Following Gamerman and Lopes (2006) consider the conditional distribution $\theta_{jk(t-1)} | \mathbf{z}_{jk}^{t-1} \sim N(m_{jk(t-1)}, C_{jk(t-1)})$, where \mathbf{z}_{jk}^{t-1} refer to the information until $t - 1$. The system equation (4.3.18) can be written as $\theta_{jkt} | \theta_{jk(t-1)}, h_{jkt} \sim N(\beta_{jkt} + \tau_{kt} h_{jkt}, \varsigma_{kt}^2)$. By properties of the normal distribution, these specifications can be combined leading to the marginal distribution:

$$\theta_{jkt} | \mathbf{z}_{jk}^{t-1}, h_{jkt} \sim N(a_{jkt}, R_{jkt}), \quad (4.3.19)$$

where

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

Thus,

$$\theta_{jkt} | \mathbf{z}_{jk}^t, h_{jkt} \sim N(m_{jkt}, C_{jkt}), \quad (4.3.20)$$

where

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

Equation 4.3.20 is referred in the literature as *Kalman Filter*. Therefore, the backward distributions are given by:

$$\theta_{jkt} | \theta_{jk(t+1)}, \mathbf{z}_{jk}^t, h_{jkt} \sim N(m_{\theta_{kt}}, C_{\theta_{kt}}), \quad (4.3.21)$$

where

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

Algorithm 4.4 FFBS algorithm

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

where $\alpha_{jk(t+1)} = \xi_{k(t+1)} - \phi_{t+1,t} \mu_{\theta_{kt}} + \tau_{kt} h_{jk(t+1)} + \sum_{r=1}^{t-1} \phi_{t+1,t} (\theta_{jkr} - \mu_{\theta_{kr}})$, for all $k = 1, \dots, K$, $j = 1, \dots, n_{kt}$ and $t = 1, \dots, T_k$. Then, a scheme to sample from the full conditional distribution of $\theta_{jk.}$ is given by algorithm 4.4.

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

$$H_{jkt} | (\cdot) \sim N \left[\frac{\tau_{kt}(\theta_{jkt} - \beta_{jtt})}{\tau_{kt}^2 + \varsigma_{kt}^2}, \frac{\varsigma_{kt}^2}{\tau_{kt}^2 + \varsigma_{kt}^2} \right] \mathbb{I}(h_{jkt} > 0). \quad (4.3.22)$$

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

To sample the correlation parameters in the case of structured correlation matrices we are considering the SVE algorithm discussed in section 3.4.3 of the Chapter 3. To illustrate this algorithm consider, for example, an ARH(1) matrix, see Table 4.1. Let $\theta_{.k.}$ being the set of all latent traits of the group k , as defined in Section 4.2 and $p(\theta_{.k.} | \mu_{\theta_k}, \phi_k, \mathbf{d}_k, \gamma_{\theta_k})$ denoting the likelihood generated by the antedependence model defined in equation (4.2.7), that is

$$\begin{aligned} p(\theta_{.k.} | \mu_{\theta_k}, \phi_k, \mathbf{d}_k, \gamma_{\theta_k}) &= \prod_{j=1}^{n_{kt}} p(\theta_{jk1} | \mu_{\theta_{k1}}, d_{k1}, \gamma_{\theta_{k1}}) \prod_{t=2}^{T_k} p(\theta_{jkt} | \mu_{\theta_{k1}}, \phi_{kt}, d_{kt}, \gamma_{\theta_{kt}}) \\ &\propto 2^{Tn_t} \prod_{j=1}^{n_{kt}} \prod_{t=1}^{T_k} \omega_{kt}^{-1} \phi_d \left(\omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt}) \right) \\ &\times \Phi_d \left[\lambda_{kt} (\omega_{kt}^{-1} (\theta_{jkt} - \beta_{jkt})) \right], \end{aligned} \quad (4.3.23)$$

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

Algorithm 4.5 present the correlation parameters sampler for the ARH(1) model for all simulation $m = 1, \dots, M$.

Algorithm 4.5 The SVE algorithm with oversampling to sample a correlation parameter considering ARH(1) matrix

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

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

1: **for** $g = 1$ to G **do**

2: Draw $\rho_{\theta_{k.g}}^{(m)} \sim p(\rho_{\theta_k})$

3: Draw $\boldsymbol{\theta}_{.k.g}^{(m)}$ from the model (4.2.7)

4: Compute $r_1(\boldsymbol{\theta}_{.k.g}^{(m)})$ the first-order sample correlation

5: **end for**

6: Choose the $\rho_{\theta_{k.g}}^{(m)}$ and $\boldsymbol{\theta}_{.k.g}^{(m)}$ whose $r_1(\boldsymbol{\theta}_{.k.g}^{(m)})$ is closest to $r_1(\boldsymbol{\theta}_{.k}^{(m-1)})$

7: Set $\rho_{\theta_k}^{(m)} = \rho_{\theta_{k.g}}^{(m)}$ and $\boldsymbol{\theta}_{.k}^{(m)} = \boldsymbol{\theta}_{.k.g}^{(m)}$ the candidate values

8: Build the ARH(1) proposed matrix $\Sigma_{\rho_{\theta_k}}^{(m)}$ using `AR1.matrix()`

9: Perform the Cholesky decomposition of $\Sigma_{\rho_{\theta_k}}^{(m)}$ to obtain the matrices $\mathbf{L}_k^{(m)}$ and $\mathbf{D}_k^{(m)}$

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

11: **if**

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

then

12: $\rho_{\theta_k}^{(m-1)} = \rho_{\theta_k}^{(m)}$

13: **end if**

For more than one correlation parameter, the algorithm 4.5 can be applied independently to each one by choosing suitable sufficient statistics. It is also possible to sample blocks of correlation parameters. This can be done by modifying line 11 of the algorithm 4.5 to allow accept/reject proposed values jointly. Also, different covariance matrices can be considered for different groups, as will see in next section. More details about the SVE are found in Appendix A.

In summary, a general algorithm to estimate the parameter's model is a combination of Gibbs sampling, FFBS, SVE and Metropolis-Hastings algorithms as we can see in algorithms 4.6 and 4.7.

Algorithm 4.6 General algorithm for unstructured matrix

- 1: Start the algorithm by choosing suitable initial values. Repeat steps 2-9.
 - 2: Simulate W_{ijkt} from $W_{ijkt}|\cdot$ for all $K = 1, \dots, K$, $i = 1, \dots, I$, $j = 1, \dots, n_{kt}$ and $t = 1, \dots, T_k$.
 - 3: Simulate Z_{ijkt} from $Z_{ijkt}|\cdot$ for all $K = 1, \dots, K$, $i = 1, \dots, I$, $j = 1, \dots, n_{kt}$ and $t = 1, \dots, T_k$.
 - 4: Simulate H_{jkt} from $H_{jkt}|\cdot$ for all $K = 1, \dots, K$, $j = 1, \dots, n_{kt}$ and $t = 1, \dots, T_k$.
 - 5: Simulate θ_{jkt} using the algorithm 4.4 for all $K = 1, \dots, K$, $j = 1, \dots, n_{kt}$ and $t = 1, \dots, T_k$.
 - 6: Simulate ζ_i from $\zeta_i|\cdot$ for all $i = 1, \dots, I$.
 - 7: Simulate c_i from $c_i|\cdot$ for all $i = 1, \dots, I$.
 - 8: Simulate α_{kl} from $\alpha_{kl}|\cdot$ using algorithm 4.1 for all $k = 1, \dots, K$ and $l = 1, \dots, 3$.
 - 9: Simulate $\sigma_{\theta_{kt}}^2$ from $\sigma_{\theta_{kt}}^2|\cdot$ using algorithm 5.2 for all $k = 1, \dots, K$ and $t = 1, \dots, T_k$.
 - 10: Simulate $\gamma_{\varepsilon_{kt}}$ from $\gamma_{\varepsilon_{kt}}|\cdot$ using algorithm 4.3 for all $k = 1, \dots, K$ and $t = 1, \dots, T_k$.
 - 11: Simulate ϕ_{tk} from $\phi_{tk}|\cdot$ for all $t = 2, \dots, T$ and $k = 1, \dots, t - 1$.
-

Algorithm 4.7 General algorithm for structured matrices

- 1: Start the algorithm by choosing suitable initial values. Repeat steps 2-9.
 - 2: Simulate W_{ijkt} from $W_{ijkt}|\cdot$ for all $K = 1, \dots, K$, $i = 1, \dots, I$, $j = 1, \dots, n_{kt}$ and $t = 1, \dots, T_k$.
 - 3: Simulate Z_{ijkt} from $Z_{ijkt}|\cdot$ for all $K = 1, \dots, K$, $i = 1, \dots, I$, $j = 1, \dots, n_{kt}$ and $t = 1, \dots, T_k$.
 - 4: Simulate H_{jkt} from $H_{jkt}|\cdot$ for all $K = 1, \dots, K$, $j = 1, \dots, n_{kt}$ and $t = 1, \dots, T_k$.
 - 5: Simulate θ_{jkt} using the algorithm 4.4 for all $K = 1, \dots, K$, $j = 1, \dots, n_{kt}$ and $t = 1, \dots, T_k$.
 - 6: Simulate ζ_i from $\zeta_i|\cdot$ for all $i = 1, \dots, I$.
 - 7: Simulate c_i from $c_i|\cdot$ for all $i = 1, \dots, I$.
 - 8: Simulate α_{kl} from $\alpha_{kl}|\cdot$ using algorithm 4.1 for all $k = 1, \dots, K$ and $l = 1, \dots, 3$.
 - 9: Simulate $\sigma_{\theta_{kt}}^2$ from $\sigma_{\theta_{kt}}^2|\cdot$ using algorithm 5.2 for all $k = 1, \dots, K$ and $t = 1, \dots, T_k$.
 - 10: Simulate $\gamma_{\varepsilon_{kt}}$ from $\gamma_{\varepsilon_{kt}}|\cdot$ using algorithm 4.3 for all $k = 1, \dots, K$ and $t = 1, \dots, T_k$.
 - 11: Simulate ρ_{θ_i} using a SVE procedure for all necessary correlation parameters.
-

4.4 Parameter recovery study

In this section we study the efficiency of our model and the proposed estimation algorithm concerning to parameter recovery. Our algorithm allows to consider different structured covariance matrices. Some examples can be seen in Table 4.1.

We considered a scenario with $K = 2$ groups with $T_1 = T_2 = 6$, that is, the same number of time-points. Responses of $n_{kt} = 500$ were simulated for $k = 1, 2$ and $t = 1, \dots, 6$ according to model (4.2.7) considering the ARH(1) matrix for the group 1 and the AD matrix for the group 2. We considered different growth curves for each group. For group 1 we considered a count curve with restriction $\mu_{\theta_1}(1) = 0$ and parameters $\alpha_{11} = \alpha_{12} = .1$, see Table 4.3. Therefore the first time-point of group 1 is considered as the reference. For group 2, we choose a mitscherlich curve with parameters $\alpha_{21} = 1$, $\alpha_{22} = 2$ and $\alpha_3 = .3$, see Table 4.2. Figure 4.1 presents the simulated mean curve for each group. Population variances were fixed as $\sigma_{\theta_1}^2 = (1.00, .77, .60, .57, .34, .40)'$ for group 1 and $\sigma_{\theta_2}^2 = (1.10, 1.15, .80, .73, .70, .65)'$. For marginal skewness coefficient we considered $\gamma_{\theta_1} = (.80, .54, .10, .12, .17, -.02)'$ for group 1 and $\gamma_{\theta_2} = (.85, .58, .19, .30, .24, -.08)'$ for group 2. For each group, the population means increase over time whereas the variances increase and then decrease. Also, we can see from Figure 4.1 that the population means of the groups are different in the first time-points and become equal along the study. For correlation parameters we have: $\rho_{\theta_1} = .8$ for the group 1 and $\rho_{\theta_2} = (.81, .89, .93, .73, .69)'$ for group 2.

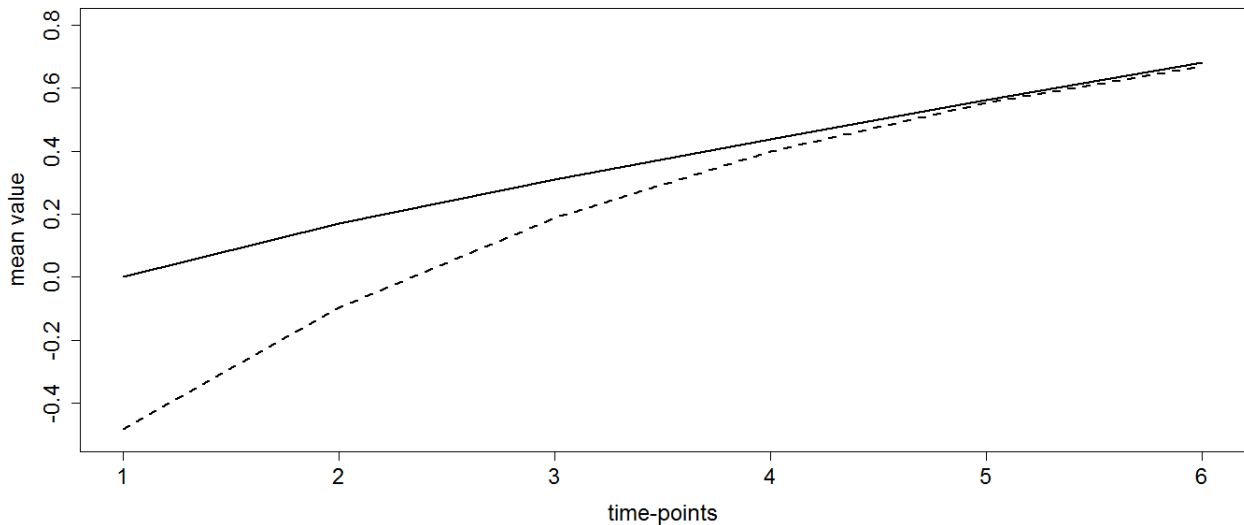


Figure 4.1: Underlying growth curves for the simulation study. Legend: Group 1 (solid line), Group 2 (dashed line)

The tests structure consists on a unique test with $I = 25$ items applied for each group along all time-points. It means that, each subject is submitted to the same items in every occasions. This design is very common in psychiatric studies and resembles the real data analyzed in next section. Table 4.4 the values for the item parameters.

Table 4.5 presents the hyperparameters for the prior distributions. The prior distribution for the population mean and variance are concentrated around zero and one, respectively. For the skewness parameter, we are assuming more probability for values near zero but allowing reasonable probabilities for the others. The discrimination parameters are assumed to vary reasonably around a satisfactory

Table 4.4: Item parameters for the recovery parameters study

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

discrimination power and for the difficulty parameter we assume a value above the mean of the reference time-point.

Table 4.5: The priors distributions hyperparameters

Hyperparameters					
μ_ζ	Ψ_ζ	$(\mu_\alpha, \sigma_\alpha^2)$	(a_σ, b_σ)	$(\mu_\phi, \sigma_\phi^2)$	$(\mu_\rho, \sigma_\rho^2)$
(1, 0)	(.5, 16)	(0, 10)	(2.1, 1.1)	(0, 10)	(0, 10)

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

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

Tables 4.6 and 4.7 present the population parameters estimates of the group 1 and 2, respectively. We can see that estimate were very close to the true values, and most of the 95% credibility intervals include the true parameters values. The estimates of some skewness and correlation parameters of the group 2 present a slight deviation of the true values. It is certainly due a random fluctuation of the generated sample. Further, Table 4.8 and Figure 4.2 present the results for the latent traits and item

parameters. In general the estimates presents small bias and small standard error, indicating that our algorithm recovered the parameters properly.

Table 4.6: Population parameters estimates of the Group 1

	True value	Mean	SD	CI(95%)
α_{11}	.100	.105	.025	[.048, .147]
α_{12}	.100	.078	.065	[-.041, .227]
$\sigma_{\theta_{11}}^2$	1.000	-	-	-
$\sigma_{\theta_{12}}^2$.770	.708	.097	[.531, .909]
$\sigma_{\theta_{13}}^2$.600	.601	.083	[.459, .787]
$\sigma_{\theta_{14}}^2$.570	.572	.074	[.438, .728]
$\sigma_{\theta_{15}}^2$.340	.354	.055	[.268, .483]
$\sigma_{\theta_{16}}^2$.400	.383	.056	[.273, .496]
$\gamma_{\theta_{11}}$.800	.738	.094	[.525, .895]
$\gamma_{\theta_{12}}$.539	.483	.217	[.046, .864]
$\gamma_{\theta_{13}}$.103	-.154	.202	[-.601, .175]
$\gamma_{\theta_{14}}$.118	.008	.188	[-.405, .373]
$\gamma_{\theta_{15}}$.168	.032	.196	[-.359, .438]
$\gamma_{\theta_{16}}$	-.022	-.043	.237	[-.594, .426]
ρ_{θ_1}	.800	.791	.014	[.761, .815]

Table 4.7: Population parameters estimates of the Group 2

	True value	Mean	SD	CI(95%)		True value	Mean	SD	CI(95%)
α_{21}	1.000	.908	.117	[.731, 1.212]	$\gamma_{\theta_{22}}$.583	.033	.185	[-.318, .438]
α_{22}	2.000	2.095	.122	[1.880, 2.373]	$\gamma_{\theta_{23}}$.193	-.682	.148	[-.914, -.318]
α_3	.300	.350	.047	[.247, .438]	$\gamma_{\theta_{24}}$.298	.067	.187	[-.298, .483]
$\sigma_{\theta_{21}}^2$	1.100	1.296	.162	[1.022, 1.632]	$\gamma_{\theta_{25}}$.241	.131	.200	[-.220, .590]
$\sigma_{\theta_{22}}^2$	1.150	1.318	.151	[1.059, 1.627]	$\gamma_{\theta_{26}}$	-.082	-.077	.194	[-.500, .306]
$\sigma_{\theta_{23}}^2$.800	.814	.099	[.618, 1.019]	$\rho_{\theta_{21}}$.810	.804	.011	[.779, .828]
$\sigma_{\theta_{24}}^2$.730	.775	.093	[.629, .972]	$\rho_{\theta_{22}}$.890	.857	.014	[.831, .881]
$\sigma_{\theta_{25}}^2$.700	.716	.090	[.531, .864]	$\rho_{\theta_{23}}$.930	.891	.022	[.845, .927]
$\sigma_{\theta_{26}}^2$.650	.675	.062	[.555, .796]	$\rho_{\theta_{24}}$.730	.782	.030	[.723, .838]
$\gamma_{\theta_{21}}$.850	.435	.171	[.074, .744]	$\rho_{\theta_{25}}$.690	.724	.052	[.622, .819]

Table 4.8: Results for the estimated latent traits and item parameters.

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

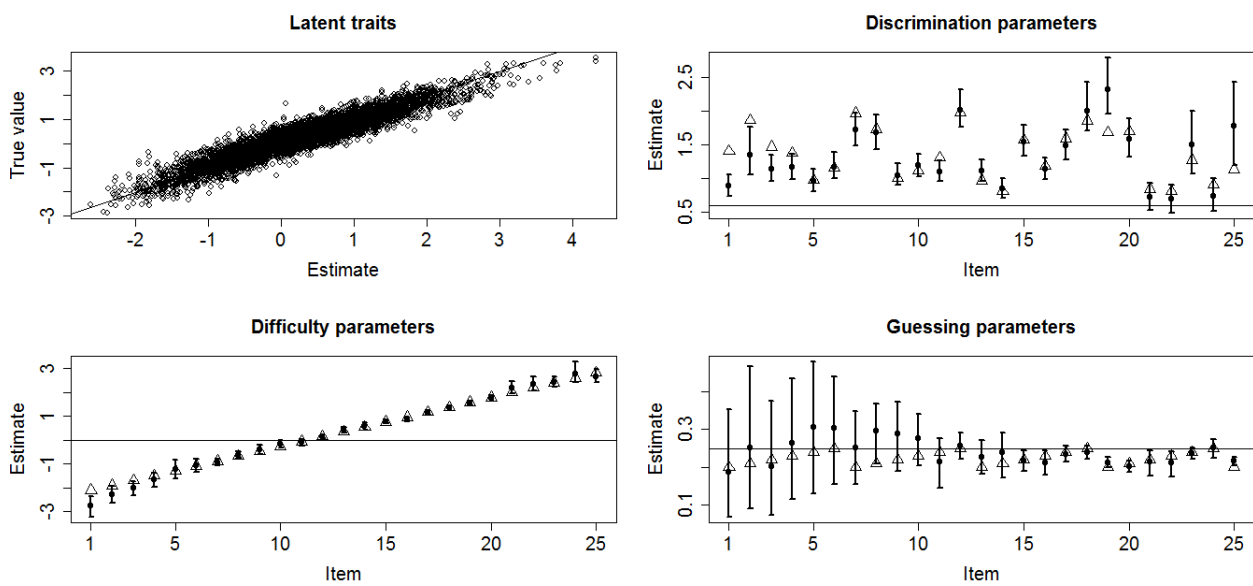


Figure 4.2: Estimates of latent traits and item parameters. Legend: circles denotes estimates, triangles denotes true values and vertical bars denote 95% credibility intervals

4.5 Real Data Analysis and Model fit Assessment

4.5.1 Model fit assessment tools

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

$$p(\mathbf{y}_t^{rep}|\mathbf{y}_t^{obs}) = \int p(\mathbf{y}_t^{rep}|\boldsymbol{\vartheta}_t)p(\boldsymbol{\vartheta}_t|\mathbf{y}_t^{obs})d_{\boldsymbol{\vartheta}_t}, \quad (4.5.1)$$

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

$$\mathbb{P}(D(\mathbf{y}_t^{rep}|\boldsymbol{\vartheta}_t) \geq D(\mathbf{y}_t^{obs}|\boldsymbol{\vartheta}_t)|\mathbf{y}_t^{obs}) = \int_{D(\mathbf{y}_t^{rep}) \geq D(\mathbf{y}_t^{obs})} p(\mathbf{y}_t^{rep}|\mathbf{y}_t^{obs})d_{\mathbf{y}_t^{rep}}, \quad (4.5.2)$$

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

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

$$D_i = \sum_l \frac{|P_{li}^O - P_{li}^E|}{P_{li}^E}, \quad (4.5.3)$$

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

4.5.2 Model Comparison

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

$$D(\vartheta) = -2\text{Log}(L(\boldsymbol{\theta}_{\cdot}, \boldsymbol{\zeta}_{\cdot}, \boldsymbol{\eta}_{\theta})P(\theta|\boldsymbol{\eta}_{\theta})). \quad (4.5.4)$$

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

$$\overline{D(\vartheta)} = \frac{1}{M} \sum_{m=1}^M D(\vartheta^{(m)}), \quad (4.5.5)$$

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

$$\begin{aligned}
\widehat{\text{DIC}} &= D(\bar{\vartheta}) + 2\rho_D, \\
\widehat{\text{EAIC}} &= \overline{D(\vartheta)} + 2\rho_D, \\
\widehat{\text{EBIC}} &= \overline{D(\vartheta)} + 2\log(n \times I),
\end{aligned}
\tag{4.5.6}$$

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

4.5.3 Amsterdam Growth and Health Longitudinal Study

Data were analyzed from the AGHLS (Amsterdam Growth and Health Longitudinal Study), a longitudinal cohort study planned to monitor the growth, health and life-style of teenagers from secondary school (Kemper and van 't Hof, 1978). The AGHLS is focused on items concerning to the relationships between anthropometry, physical activity, cardiovascular disease risk, life-style, musculoskeletal health, psychological health and wellbeing. The sample considered corresponds to 452 participants who were followed over the period of 1990-2006 with a maximum of four assessment conditions (1991, 1993, 2000 and 2006) for each subject. The measurement years will be referred as 1, 2, 3 and 4, respectively. Two groups ($k = 2$) were considered, male students and female students. The frequency of students in each time-point according to gender is presented in Table 4.9. A Dutch version of the Spielberger state trait anxiety inventory (STAI-DY) (Spielberger, 1989) was used to measure the trait "state anxiety", using a total of thirteen items with four response categories. High scores on this test are related to high anxiety.

Table 4.9: Number of subjects according to gender

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

A explanatory analysis was conducted through the symmetric multiple group IRT model (Azevedo et al., 2016). This model assumes that subjects were nested in groups and latent traits are independently distributed over groups. Then we considered the measurement years as four groups. Table 4.10 presents the Parson's correlations estimated for the complete pairs of estimated latent traits corresponding to years 1 to 4 according by gender. The results suggests that the latent traits are correlated over time-points. Figure 4.3 presents the estimated population means for each gender. The results indicate a difference among population means over the years and between groups, being the female's means higher than the male's ones.

Some multiple group longitudinal skew IRT models are compared by using the statistics of model comparison discussed before. Firstly, we considered the count growth curve for both groups and compare two correlation structures: AD and HT. These matrices seem to be the most appropriated ones, since the time-points are not equally spaced. We used the same correlation matrix for both groups, for simplicity. The results are presented in Table 4.11. The AD model outperforms the HT according to all statistics. Figures 4.4 and 4.5 present the estimated growth curves considering the AD matrix and the estimated means with 95% central credibility intervals. The results are in agreement with the explanatory analysis. Table 4.12 presents the results of the comparison of the curves. The count was selected according to all comparison statistics. Therefore, the selected model considers a count growth curve and an AD covariance matrix for the two groups.

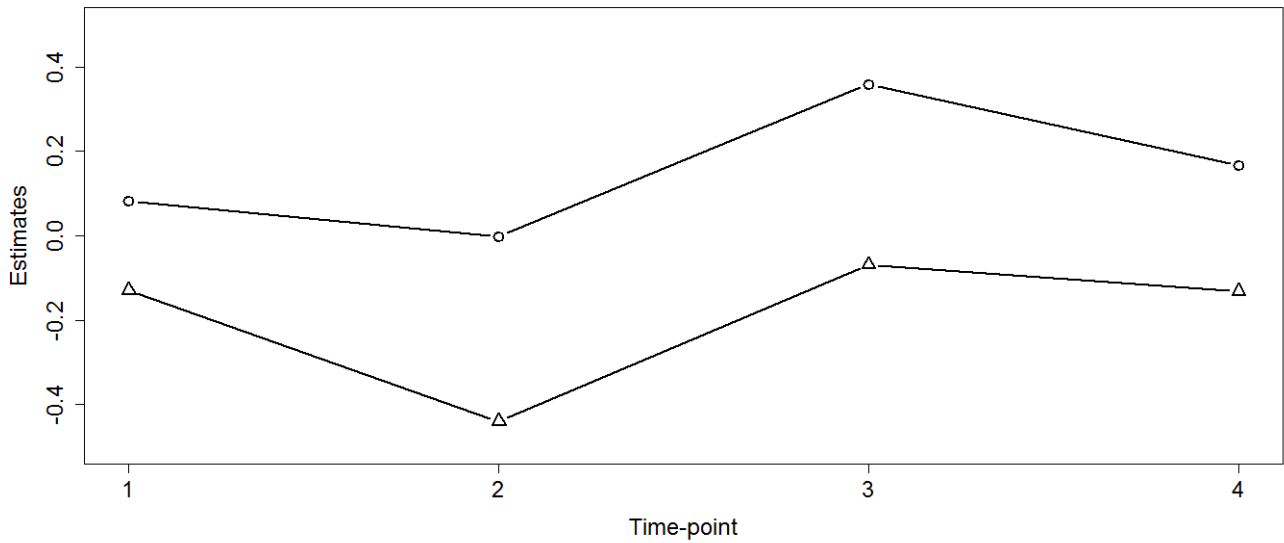


Figure 4.3: Estimates of the population means considering the Multiple Group Model. Legend: Female (— o —), Male (— Δ —)

Table 4.10: Estimated correlation structure for female and male group

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

Table 4.11: Statistics for comparison of the correlation structures

	DIC	EAIC	EBIC
AD	12960.41	13795.53	20113.33
HT	14423.47	15588.61	24403.07

Table 4.12: Statistics for comparison of the growth curves

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

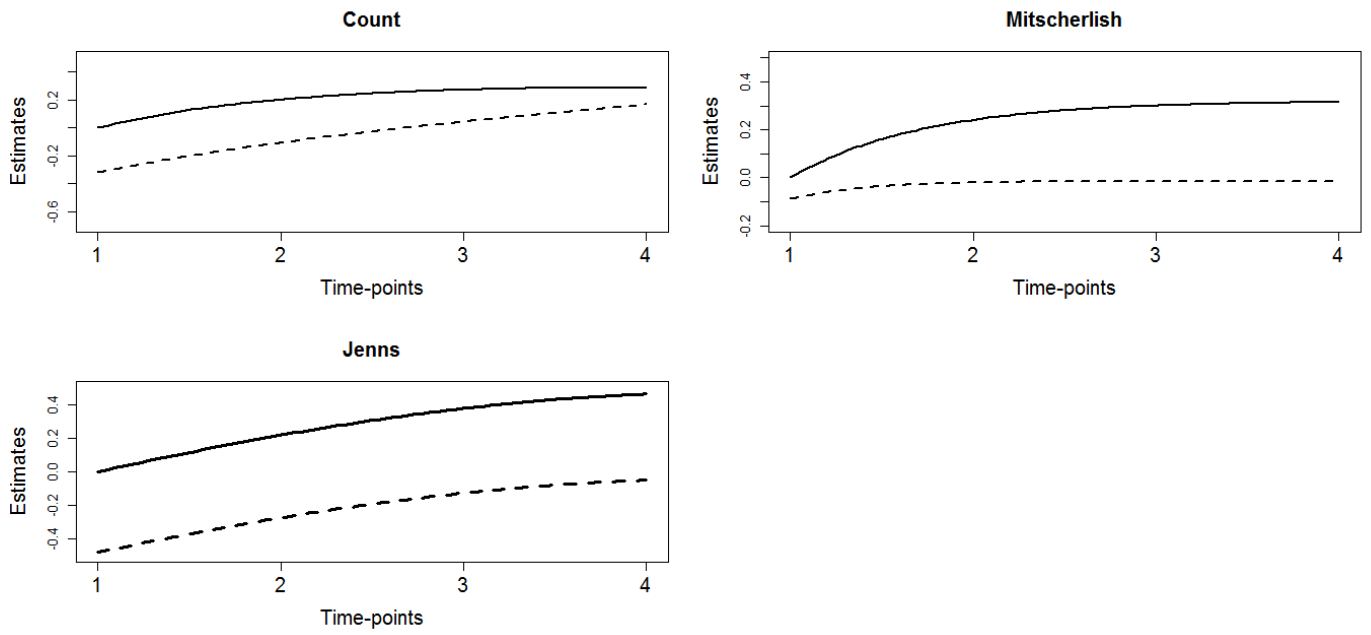


Figure 4.4: Estimated growth curves. Legend: Female (solid line), Male (dashed line)

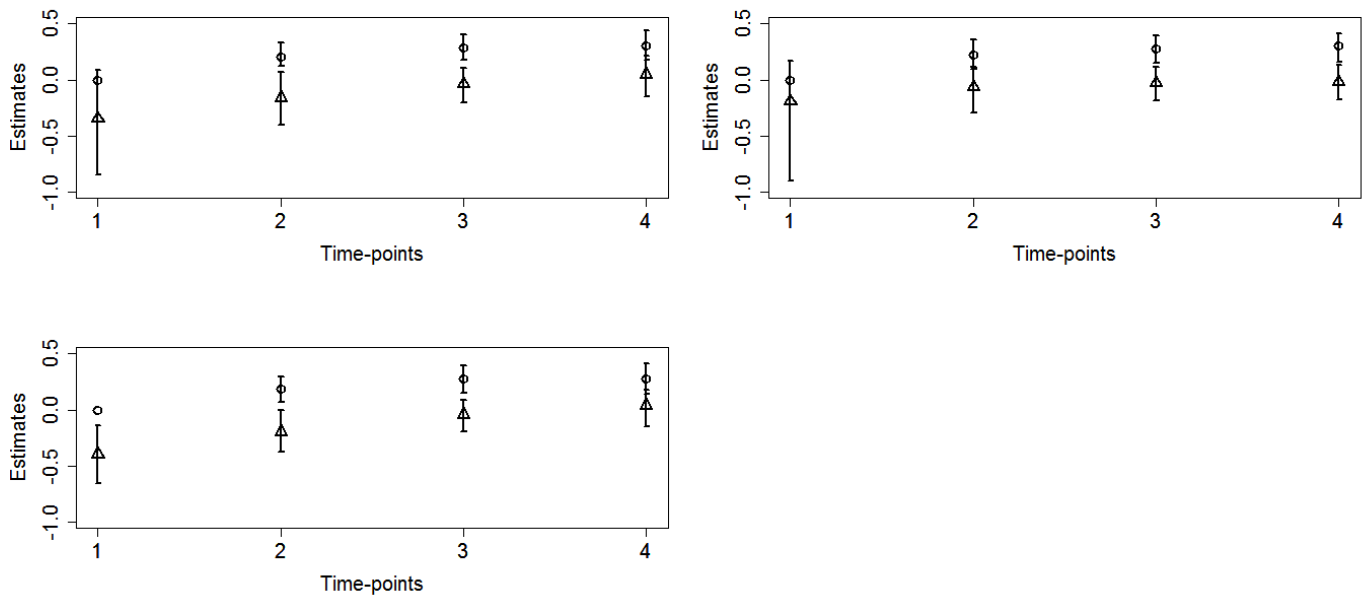


Figure 4.5: Estimated means with 95% central credibility intervals. Legend: Female ($- \circ -$), Male ($- \triangle -$).

Figures 4.6 and 4.7 present the observed and predicted scores with 95% credibility intervals. We can see that most of the observed scores distribution are well within the intervals. Some discrepancy is observed considering the Year 2000 of the male group. However, in general the results are indicating that the model is well fitted. Figures 4.8 and 4.9 present smoothed histograms of the latent trait estimates according to the skew AD model, with theoretical curves. The plot of the theoretical curves of the skew AD model was based on latent traits simulated via model 4.2.7. The estimated latent trait distributions are heavy-tailed and multimodal in some cases. This behavior is probably due to the few number of observations available. However, our model provided a good description of the latent trait distributions behavior. The population parameters estimates can be seen in Tables 4.13 and 4.14. These Tables also present the estimates of $\mathbb{P}(\gamma_{\theta_{kt}} \notin [-.13, .13])$, denoted by \hat{p}_{γ} . As discussed in Chapter 3, values of the skewness coefficient in the interval $[-.13, .13]$, indicate symmetric latent trait distribution. We considered that the asymmetry coefficient of the latent traits distribution is not null when we had more than 60% of the sampled asymmetry coefficient values outside the interval $[-.13, .13]$. Therefore, only time-points 1 and 3 of the female group can be consider asymmetric.

The correlation parameters estimates were high and statistically significant. Figure 4.10 presents the differences between the group means over the measurement years. We can see that mean of the groups are statistically different, except in time-point 1. Figure 4.11 presents the Bayesian p -values for the item parameters. The results indicate that all items are well fitted. The estimates of the item parameters are presented in Figure 4.12. In general items presented a good discrimination (estimates greater than .6) and high difficulty. The guessing parameter estimates were low and concentrated around .05, indicating that the two parameters probit model could be preferable.

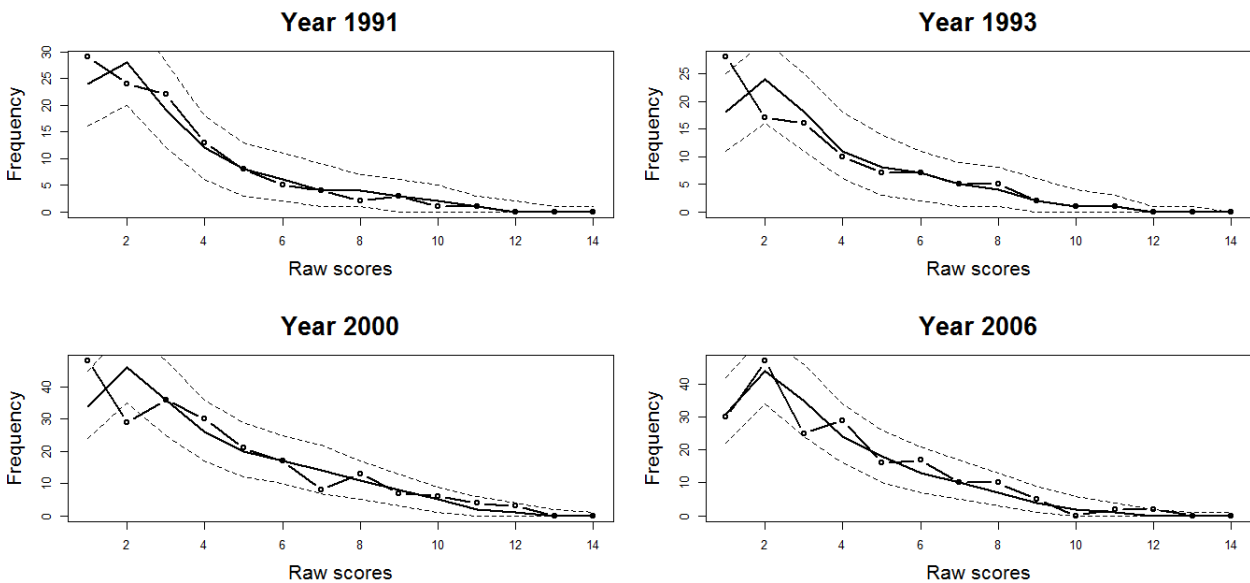


Figure 4.6: Observed and predicted scores distributions with 95% credibility intervals for female students

Table 4.13: Estimates of the population parameters for the female students

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

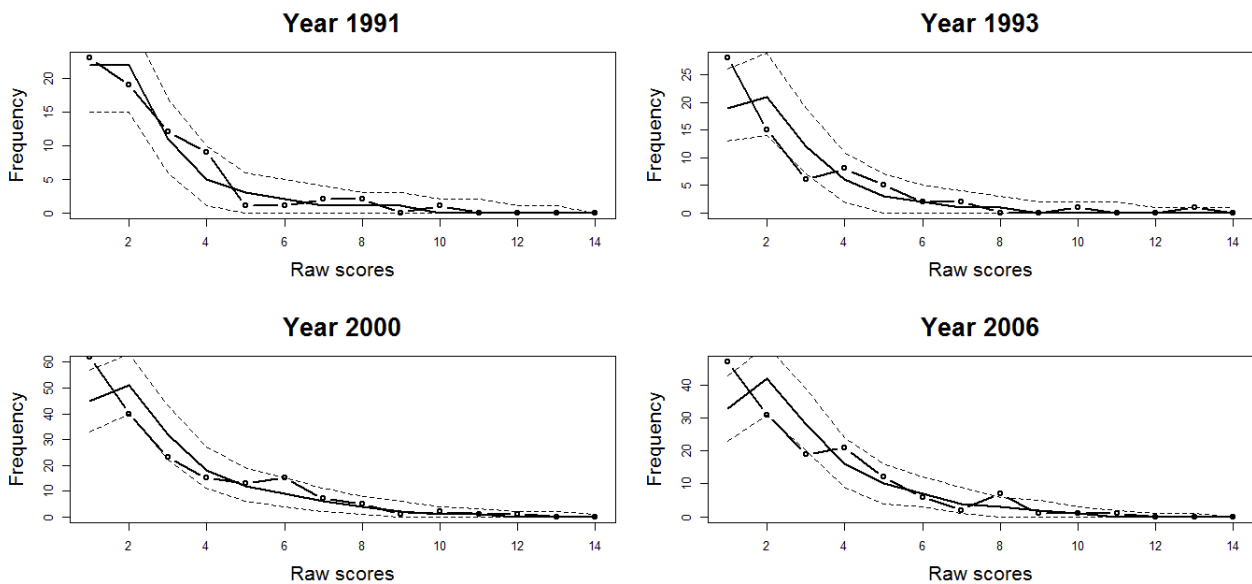


Figure 4.7: Observed and predicted scores distributions with 95% credibility intervals for male students

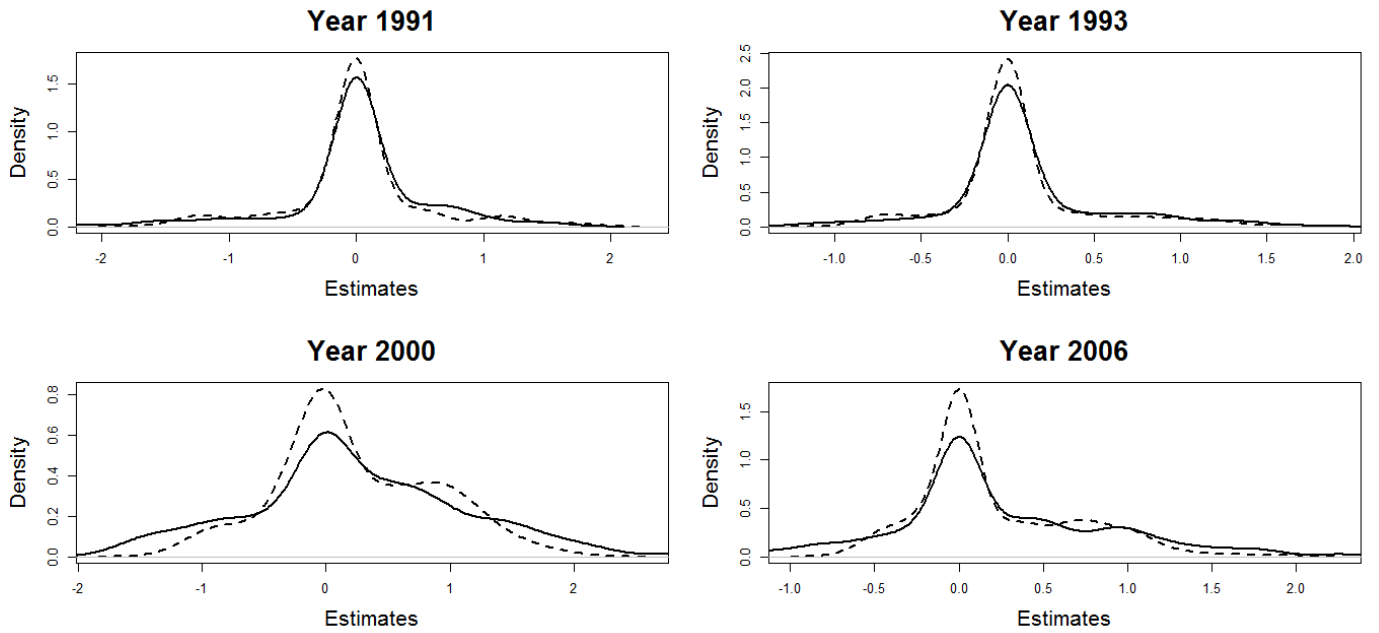


Figure 4.8: Estimated latent trait distributions for female students. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)

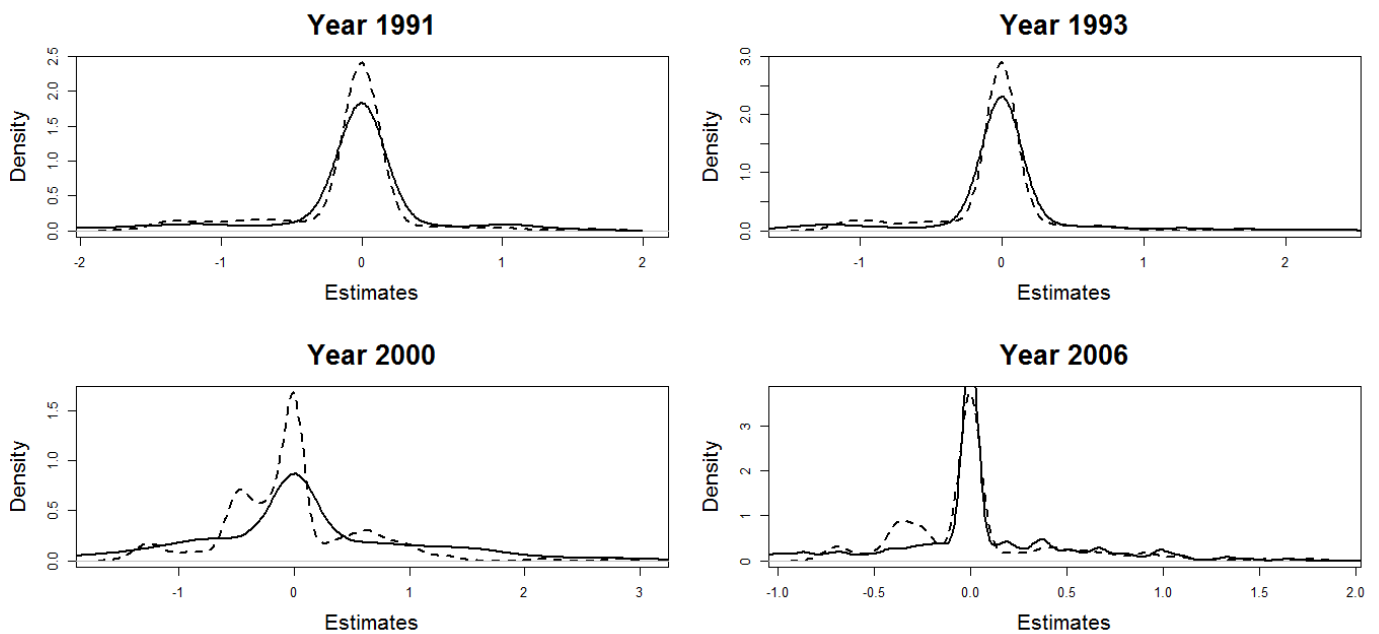


Figure 4.9: Estimated latent trait distributions for male students. Legend: Smoothed histograms (dashed line), Theoretical curve (Solid line)

Table 4.14: Estimates of the population parameters for the male students

	Mean	SD	CI(95%)	\hat{p}_γ
$\alpha_{\theta_{12}}$	-.370	.231	[-.815, .135]	–
$\alpha_{\theta_{22}}$.297	.386	[-.224, 1.296]	–
α_{θ_3}	.001	.144	[-.307, .174]	–
$\sigma_{\theta_{12}}^2$.997	.534	[.278, 2.354]	–
$\sigma_{\theta_{22}}^2$.730	.425	[.229, 1.760]	–
$\sigma_{\theta_{32}}^2$	1.327	.471	[.713, 2.476]	–
$\sigma_{\theta_{42}}^2$.540	.274	[.236, 1.181]	–
$\gamma_{\theta_{12}}$	-.057	.272	[-.626, .508]	.576
$\gamma_{\theta_{22}}$	-.001	.265	[-.572, .558]	.531
$\gamma_{\theta_{32}}$.019	.237	[-.446, .524]	.504
$\gamma_{\theta_{42}}$	-.108	.264	[-.660, .412]	.557
$\rho_{\theta_{12}}$.842	.041	[.731, .890]	–
$\rho_{\theta_{22}}$.794	.062	[.655, .877]	–
$\rho_{\theta_{32}}$.791	.112	[.594, .971]	–

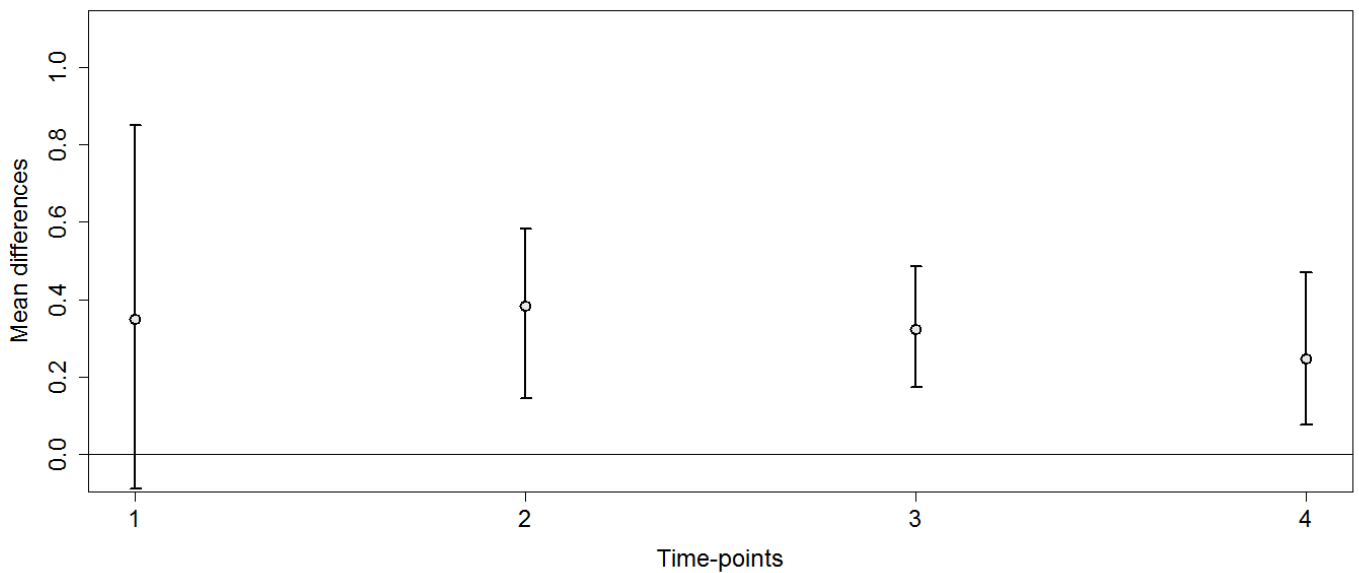


Figure 4.10: Estimated difference between the group means along the years. Legend: Estimate (circles), 95% credibility intervals (vertical bars)

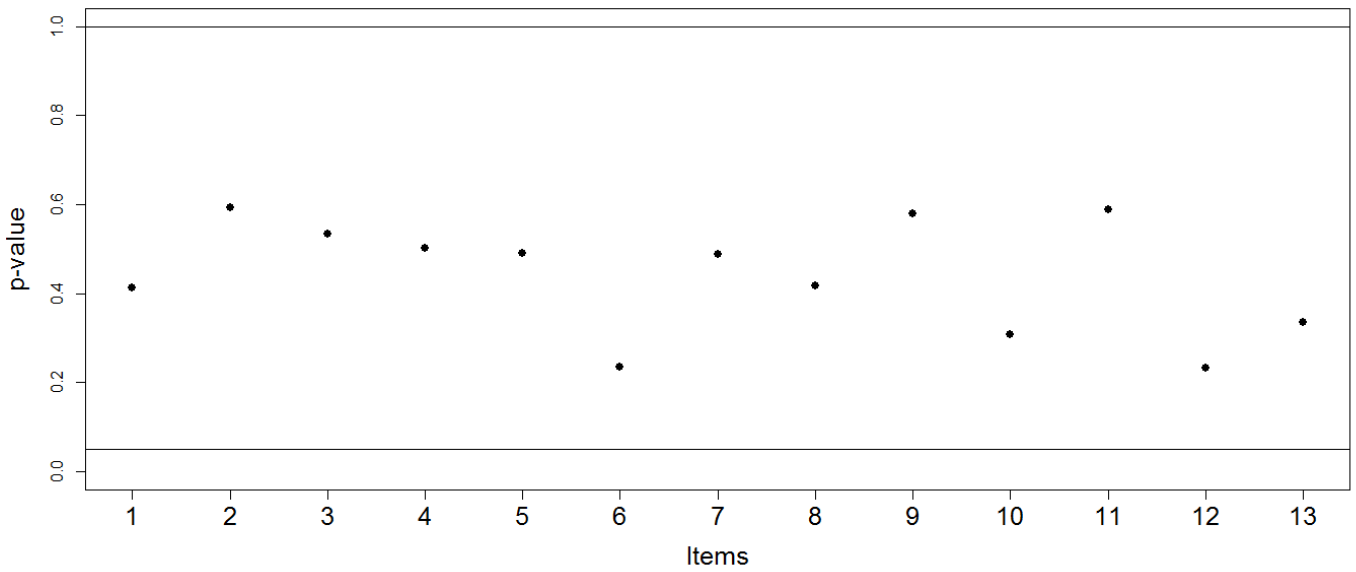
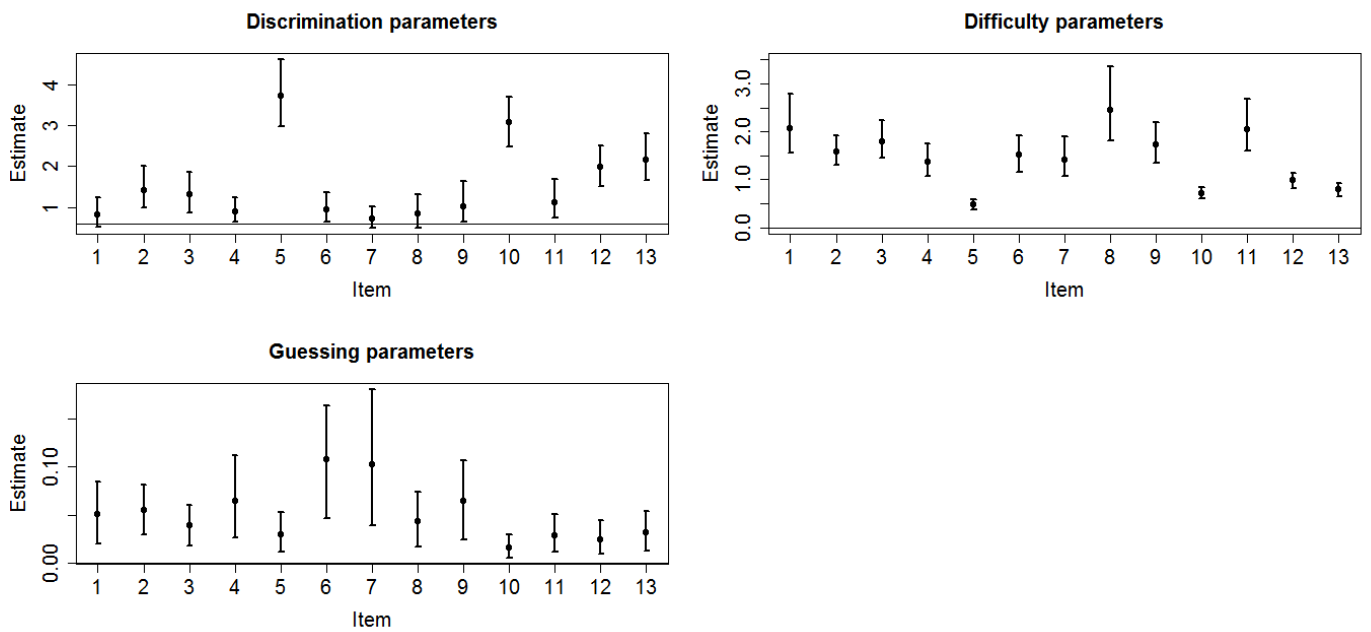
Figure 4.11: Bayesian p -values for item parameters

Figure 4.12: Item parameters estimates. Legend: estimates (circles), 95% credibility intervals (vertical bars)

4.6 Concluding Remarks

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

Chapter 5

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

Abstract

In this chapter we introduce longitudinal IRT model considering skewed latent traits distribution, based on a Gaussian copula function. Differently of the antedependence approach considered in previous chapters, 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.

5.1 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 chapter 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, presented in the previous chapters. A clear advantage of the copula approach with respect to the antedependence models is its great flexibility to define the marginal latent trait distributions. As we saw in Chapter 3, 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 chapter is outlined as follows. In section 5.2 we introduce the model and present their main properties. In section 5.3 we describe the MCMC algorithm developed to parameter estimation. In section 5.4 a comparison between the copula and antedependence approaches are performed by using simulated data. In section 5.5 a real data set is analyzed and the copula and antedependence models are compared. Finally, in section 5.6 we presented some comments and conclusions.

5.2 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

$$F_{\mathbf{X}}(x_1, \dots, x_T) = C(F_{X_1}(x_1), \dots, F_{X_T}(x_T)). \quad (5.2.1)$$

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

$$f(x_1, \dots, x_T) = c(F_1(x_1), \dots, F_T(x_T)) \prod_{t=1}^T f_t(x_t). \quad (5.2.2)$$

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

Skew longitudinal IRT model via copulas

Following the notation defined in the Chapter 2, 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,

$$\begin{aligned} Y_{ijt} | \theta_{jt}, \zeta_i &\sim \text{Bernoulli}(P_{ijt}), \\ P_{ijt} = P(Y_{ijt} | \theta_{jt}, \zeta_i) &= c_i + (1 - c_i) \Phi(a_i \theta_{jt} - b_i). \end{aligned} \quad (5.2.3)$$

$$\boldsymbol{\theta}_j | \boldsymbol{\eta}_\theta \sim G_T(\boldsymbol{\mu}_\theta, \boldsymbol{\sigma}_\theta, \boldsymbol{\gamma}_\theta, \boldsymbol{\varphi}), \quad (5.2.4)$$

where G_T denotes a general T -dimensional distribution, whose margins are $CSN(\mu_{\theta_t}, \sigma_{\theta_t}^2, \gamma_{\theta_t})$, that is, centred skew-normal distributions according to the notation defined in Chapter 3, where the joint density is given by

$$g(\theta_{j1}, \dots, \theta_{jT}; \boldsymbol{\varphi}) = c(F_{\theta_{j1}}, \dots, F_{\theta_{jT}}; \boldsymbol{\varphi}) \prod_{t=1}^T f_{\theta_{jt}}, \quad (5.2.5)$$

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 .

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 decreases 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 exploit the time ordering of the margins, and it may to lead to more accurate estimates, see Smith et al. (2010). In this work we will consider Gaussian copulas.

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}_θ and $\Phi(\cdot)$ be the distribution function of a unidimensional standard normal distribution. The Gaussian copula function for a subject j is defined by

$$C(u_{j1}, \dots, u_{jT}; \mathbf{R}_\theta) = \Phi_T\left(\Phi^{-1}(u_{j1}), \dots, \Phi^{-1}(u_{jT}); \mathbf{R}_\theta\right). \quad (5.2.6)$$

Derivation of the copula density is straightforward by differentiation of the equation (5.2.6), so that

$$c(u_{j1}, \dots, u_{jT}; \boldsymbol{\eta}_\theta, \mathbf{R}_\theta) = \frac{\partial}{\partial \mathbf{u}_j} C(\mathbf{u}_j) = |\mathbf{R}_\theta|^{-1/2} \exp\left\{-\frac{1}{2} \mathbf{x}'_j \mathbf{R}_\theta^{-1} \mathbf{x}_j\right\}. \quad (5.2.7)$$

Therefore, according to the equations (5.2.5) and (5.2.7) the joint probability density of the latent traits for a subject j is given by

$$p(\boldsymbol{\theta}_j | \boldsymbol{\eta}_\theta, \mathbf{R}_\theta) = |\mathbf{R}_\theta|^{-1/2} \exp\left\{-\frac{1}{2} \mathbf{x}'_j \mathbf{R}_\theta^{-1} \mathbf{x}_j\right\} \prod_{t=1}^T f_t(\theta_{jt}; \boldsymbol{\eta}_{\theta_t}), \quad (5.2.8)$$

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}_θ will be assumed to be a structured correlation matrix. Table 5.1 presents four possible structures of \mathbf{R}_θ . 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. Following the notation of the Chapter 3, the centered skew-normal density is given by:

$$f_t(\theta_{jt} | \boldsymbol{\eta}_{\theta_t}) = 2\omega_t^{-1} \phi\left(\omega_t^{-1}(\theta_{jt} - \xi_t)\right) \Phi\left[\lambda_t(\omega_t^{-1}(\theta_{jt} - \xi_t))\right], \quad (5.2.9)$$

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

$$\begin{aligned} \xi_t &= \mu_{\theta_t} - \sigma_{\theta_t} \gamma_{\theta_t}^{1/3} s, \\ \omega_t &= \sigma_{\theta_t} \sqrt{1 + \gamma_{\theta_t}^{2/3} s^2}, \\ \lambda_t &= \frac{\gamma_{\theta_t}^{1/3} s}{\sqrt{r^2 + s^2 \gamma_{\theta_t}^{2/3} (r^2 - 1)}}, \text{ where} \\ s &= \left(\frac{2}{4 - \pi}\right)^{1/3}. \end{aligned} \quad (5.2.10)$$

Table 5.1: 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 AD	$\begin{pmatrix} 1 & \rho_{\theta_1} & \dots & \prod_{t=1}^{T-1} \rho_{\theta_t} \\ \rho_{\theta_1} & 1 & \dots & \prod_{t=2}^{T-1} \rho_{\theta_t} \\ \vdots & \vdots & \ddots & \vdots \\ \prod_{t=1}^{T-1} \rho_{\theta_t} & \prod_{t=2}^{T-1} \rho_{\theta_t} & \dots & 1 \end{pmatrix}$

5.3 Bayesian Estimation and MCMC Algorithms

As made in the previous Chapters, we will consider the augmented data scheme to implement the MCMC algorithm. Essentially, we intend to obtain as many as possible of full 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

$$\begin{aligned} \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 1, \theta_{jt}, \zeta_i) &\propto \Phi(a_i \theta_{jt} - b_i) \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 1, \theta_{jt}, \zeta_i) &\propto c_i (1 - \Phi(a_i \theta_{jt} - b_i)) \\ \mathbb{P}(W_{ijt} = 1 | Y_{ijt} = 0, \theta_{jt}, \zeta_i) &= 0 \\ \mathbb{P}(W_{ijt} = 0 | Y_{ijt} = 0, \theta_{jt}, \zeta_i) &= 1. \end{aligned} \tag{5.3.1}$$

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

$$Z_{ijt} | (\theta_{jt}, \zeta_i, w_{ijt}) = \begin{cases} N(a_i \theta_{jt} - b_i, 1) \mathbb{I}_{(z_{ijt} \geq 0)}, & \text{if } w_{ijt} = 1, \\ N(a_i \theta_{jt} - b_i, 1) \mathbb{I}_{(z_{ijt} < 0)}, & \text{if } w_{ijt} = 0. \end{cases} \quad (5.3.2)$$

The original response can be represented by

$$Y_{ijt} = \mathbb{I}(Z_{ijt} > 0) \mathbb{I}(W_{ijt} = 1) + \mathbb{I}(Z_{ijt} \leq 0) \mathbb{I}(W_{ijt} = 0), \quad (5.3.3)$$

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

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

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

$$\begin{aligned} L(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta}, |\mathbf{z}_{..}, \mathbf{w}_{..}, \mathbf{y}_{..}, \mathbf{I}) &\propto \prod_{t=1}^T \prod_{j=1}^{n_t} \prod_{i \in I_{jt}} \exp \left\{ -0.5(z_{ijt} - a_i \theta_{jt} + b_i)^2 \right\} \mathbb{I}_{(z_{ijt}, w_{ijt})} \\ &\times p(w_{ijt} | y_{ijt}, \theta_{jt}, \zeta_i), \end{aligned} \quad (5.3.4)$$

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 $p(w_{ijt} | y_{ijt}, \theta_{jt}, \zeta_i)$ follows from (5.3.1).

5.3.1 Prior and posterior distributions

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

$$p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_{\theta} | \boldsymbol{\eta}_{\zeta}, \boldsymbol{\eta}_{\eta}) = \left\{ \prod_{t=1}^T \prod_{j=1}^{n_t} p(\theta_{jt} | \boldsymbol{\eta}_{\theta_t}) \right\} \left\{ \prod_{i=1}^I p(\zeta_i | \boldsymbol{\eta}_{\zeta}) \right\} \left\{ \prod_{t=1}^T p(\boldsymbol{\eta}_{\theta_t} | \boldsymbol{\eta}_{\eta}) \right\}, \quad (5.3.5)$$

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 (5.2.9). For the item parameters we have:

$$p(\boldsymbol{\zeta}_{i(-c_i)}) \propto \exp \left[-0.5(\boldsymbol{\zeta}_{i(-c_i)} - \boldsymbol{\mu}_{\zeta})' \boldsymbol{\Psi}_{\zeta}^{-1} (\boldsymbol{\zeta}_{i(-c_i)} - \boldsymbol{\mu}_{\zeta}) \right] \mathbb{I}_{(a_i > 0)} \quad (5.3.6)$$

and

$$c_i \sim \text{Beta}(a_c, b_c), \quad (5.3.7)$$

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

$$\begin{aligned} p(\mu_{\theta_t}, \sigma_{\theta_t}^2, \gamma_{\theta_t}, \rho_{\theta_t}) &= p(\mu_{\theta_t})p(\sigma_{\theta_t}^2)p(\gamma_{\theta_t})p(\rho_{\theta_t}) \\ &= p(\mu_{\theta_t})p(\sigma_{\theta_t}^2)p(\gamma_{\theta_t}) \prod_{t=1}^{T-1} p(\rho_{\theta_t}) \end{aligned} \quad (5.3.8)$$

$$\begin{aligned} \mu_{\theta_t} &\sim N(m_\mu, \sigma_\mu^2), \\ \sigma_{\theta_t}^2 &\sim IG(a_\sigma, b_\sigma), \\ \gamma_{\theta_t} &\sim N(\mu_\gamma, \sigma_\gamma^2) \mathbb{I}_{[-.99527, .99527]}, \quad t = 1, \dots, T, \\ \rho_{\theta_t} &\sim N(\mu_\rho, \sigma_\rho^2) \mathbb{I}_{(0,1)}, \end{aligned} \quad (5.3.9)$$

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 (5.3.4) and the prior distribution in equations (5.2.8), (5.2.9), (5.3.6), (5.3.7) and (5.3.9), the joint posterior distribution is given by:

$$\begin{aligned} p(\boldsymbol{\theta}_{..}, \boldsymbol{\zeta}, \boldsymbol{\eta}_\theta, |\mathbf{z}_{..}, \mathbf{w}_{..}, \mathbf{y}_{..}) &\propto \left\{ \prod_{t=1}^T \prod_{j=1}^n \prod_{i \in I_{jt}} \exp \left\{ -.5(z_{ijt} - a_i \theta_{jt} + b_i)^2 \right\} p(w_{ijt} | y_{ijt}, \theta_{jt}, \zeta_i) \mathbb{I}_{(z_{ijt}, w_{ijt})} \right\} \\ &\times \left\{ \prod_{j=1}^{n_t} p(\boldsymbol{\theta}_j | \boldsymbol{\eta}_{\theta_t}, \mathbf{R}_\theta) \right\} \left\{ \prod_{i=1}^I p(\zeta_i | \boldsymbol{\eta}_\zeta) \right\} \\ &\times \left\{ \prod_{t=1}^T p(\boldsymbol{\eta}_{\theta_t} | \boldsymbol{\eta}_\eta) \right\}. \end{aligned} \quad (5.3.10)$$

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,

$$\zeta_{i(-c_i)} | (\cdot) \sim N(\widehat{\Psi}_{\zeta_i} \widehat{\zeta}_i, \widehat{\Psi}_{\zeta_i}), \quad (5.3.11)$$

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

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,

$$c_i | (\cdot) \sim \text{Beta}(s_i + a_c - 1, t_i - s_i + b_c - 1), \quad (5.3.12)$$

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:

$$\begin{aligned} p(\mu_{\theta_t}^{(m)} | \mu_{\theta_t}^{(m-1)}) &= N(\mu_{\theta_t}^{(m-1)}, \sigma_0^2), \\ p((\sigma_{\theta_t}^2)^{(m)} | (\sigma_{\theta_t}^2)^{(m-1)}) &= \text{Lognormal}((\sigma_{\theta_t}^2)^{(m-1)}, \sigma_0^2), \\ p(\gamma_{\theta_t}^{(m)} | \gamma_{\theta_t}^{(m-1)}) &= U(\nu_1(\gamma_{\theta_t}^{(m-1)}), \nu_2(\gamma_{\theta_t}^{(m-1)})) \text{ for all } t = 1, \dots, T, \end{aligned} \quad (5.3.13)$$

where

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

The constant Δ_γ is previously defined, see Azevedo et al. (2011). The superscript (m) indicates the estimate at the iteration m of the MCMC algorithm. Algorithms 5.1 to 5.3 correspond to the Metropolis-Hastings steps to simulate population mean, variance and skewness coefficient, respectively, for all $t = 1, \dots, T$.

Algorithm 5.1 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 5.2 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 5.3 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

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

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

5.3.2 Correlation parameters sampler

Concerning the correlation parameters, we have that the full conditional distribution of ρ_θ is given by:

$$\begin{aligned} p(\rho_\theta | (\cdot)) &\propto |\mathbf{R}_\theta|^{-1/2} \left(\prod_{j=1}^{n_t} \exp \left\{ -\frac{1}{2} \mathbf{x}_j' \mathbf{R}_\theta^{-1} \mathbf{x}_j \right\} \right) p(\rho_\theta) \\ &= \prod_{j=1}^{n_t} p(\mathbf{x}_j | \boldsymbol{\eta}_{\theta_t}) p(\rho_\theta). \end{aligned} \quad (5.3.14)$$

We can see that, random variables $\mathbf{X}_j \sim N_T(\mathbf{0}, \mathbf{R}_\theta)$ for all $t = 1, \dots, T$. Therefore, an antedependence model, in the sense of the Chapter 2 can be considered. The distribution of \mathbf{X}_j can be expressed as:

$$X_{jt} = \sum_{k=1}^{t-1} \phi_{tk} X_{jk} + \varepsilon_{jt}, \quad \varepsilon_{jt} \stackrel{i.i.d.}{\sim} N(0, d_t), \quad t = 1, 2, \dots, T. \quad (5.3.15)$$

Equivalently,

$$\boldsymbol{\varepsilon}_j = \mathbf{L}_X \mathbf{X}_j, \quad \boldsymbol{\varepsilon}_j \stackrel{i.i.d.}{\sim} N_T(\mathbf{0}, \mathbf{D}_X), \quad (5.3.16)$$

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 ϕ_{tk} are the so called autoregressive generalized parameters. According to the result (2.2.6) of the Chapter 2 we have:

$$\mathbf{L}_X \mathbf{R}_\theta \mathbf{L}_X' = \mathbf{D}_X. \quad (5.3.17)$$

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 ρ_θ . This approach allow us to consider an adaptation of the SVE algorithm discussed in subsection 2.3.3 of the chapter 2. Algorithm 5.4 describes the correlation parameter sampler considering the AR(1) matrix, for example.

Notation $p(\mathbf{x} | \rho_\theta)$ represents the likelihood generated by the model (5.3.15). 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 5.4 can be applied for each parameter independently. More details about the SVE are found in Appendix A.

Algorithm 5.4 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 (5.3.15)

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**

5.3.3 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 $\boldsymbol{\theta}_j$, 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 5.5 describes the latent traits sampler.

Notation $p(Y|\boldsymbol{\theta}_j)$ denotes the likelihood generated by the model (5.2.3). That is,

$$p(Y|\boldsymbol{\theta}_j, \boldsymbol{\zeta}) = \prod_{t=1}^T \prod_{j=1}^{n_t} \prod_{i \in \mathcal{I}_{jt}} P_{ijt}^{y_{ijt}} (1 - P_{ijt})^{1-y_{ijt}}, \quad (5.3.18)$$

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 5.6.

5.4 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 in Chapter 3. For short, we will refer these models as copula and Cholesky models, respectively. Our objective 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

Algorithm 5.5 The SVE algorithm with oversampling to sample latent traits

- 1: **for** $q = 1$ to Q **do**
 - 2: Draw $\boldsymbol{\theta}_{j,q}^{(m)}$ based on definition (5.2.6)
 - 3: Draw $Y_q^{(m)}$ from the model (5.2.3)
 - 4: Compute $s(Y_q^{(m)})$ the test score
 - 5: **end for**
 - 6: Choose the $\boldsymbol{\theta}_{j,q}^{(m)}$ whose $s(Y_q^{(m)})$ is closest to $s(Y)$
 - 7: Set $\boldsymbol{\theta}_j^{(m)} = \boldsymbol{\theta}_{j,q}^{(m)}$ and $Y^{(m)} = Y_q^{(m)}$
 - 8: Draw $u \sim U(0, 1)$
 - 9: **if**

$$\left(u < \min \left\{ 1, \frac{p(Y|\boldsymbol{\theta}_j^{(m)}, \boldsymbol{\zeta}^{(m-1)})p(Y^{(m)}|\boldsymbol{\theta}_j^{(m-1)}, \boldsymbol{\zeta}^{(m-1)})}{p(Y|\boldsymbol{\theta}_j^{(m-1)}, \boldsymbol{\zeta}^{(m-1)})p(Y^{(m)}|\boldsymbol{\theta}_j^{(m)}, \boldsymbol{\zeta}^{(m-1)})} \right\} \right)$$
 - then**
 - 10: $\boldsymbol{\theta}_j^{(m-1)} = \boldsymbol{\theta}_j^{(m)}$
 - 11: **end if**
-

Algorithm 5.6 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 θ_{jt} using the algorithm 5.5 for all $j = 1, \dots, n_t$ e $t = 1, \dots, T$.
 - 5: Simulate $\boldsymbol{\zeta}_i$ from $\boldsymbol{\zeta}_i|(\cdot)$ for all $i = 1, \dots, I$.
 - 6: Simulate c_i from $c_i|(\cdot)$ for all $i = 1, \dots, I$.
 - 7: Simulate μ_{θ_t} using the algorithm 5.1 for all $t = 1, \dots, T$.
 - 8: Simulate $\sigma_{\theta_t}^2$ using the algorithm 5.2 for all $t = 1, \dots, T$.
 - 9: Simulate γ_{θ_t} using the algorithm 5.3 for all $t = 1, \dots, T$.
 - 10: Simulate ρ_{θ_t} by using SVE algorithm for all $t = 1, \dots, T - 1$.
-

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 (5.2.3). 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) (see Table B.2 in appendix). 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 5.2.8 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. (2012b). 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 5.2 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 5.2: The priors distributions hyperparameters

Hyperparameters					
$\boldsymbol{\mu}_\zeta$	$\boldsymbol{\Psi}_\zeta$	$(m_\mu; \sigma_\mu^2)$	(a_σ, b_σ)	$(\mu_\rho, \sigma_\rho^2)$	$(\mu_\gamma, \sigma_\gamma^2)$
(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 ϑ 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: $\frac{1}{n_p} \sum_{l=1}^{n_p} (\hat{\vartheta}_l - \bar{\hat{\vartheta}})^2$ 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 5.3 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 correlations (see equation (3.3.13) of the Chapter 3), unlike the copula model. Furthermore, the copula model recovered properly all parameters. Figure 5.1 presents the estimated latent trait distribution with theoretical curves of the models. Figure 5.1 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 5.4 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 5.2 and 5.3 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.

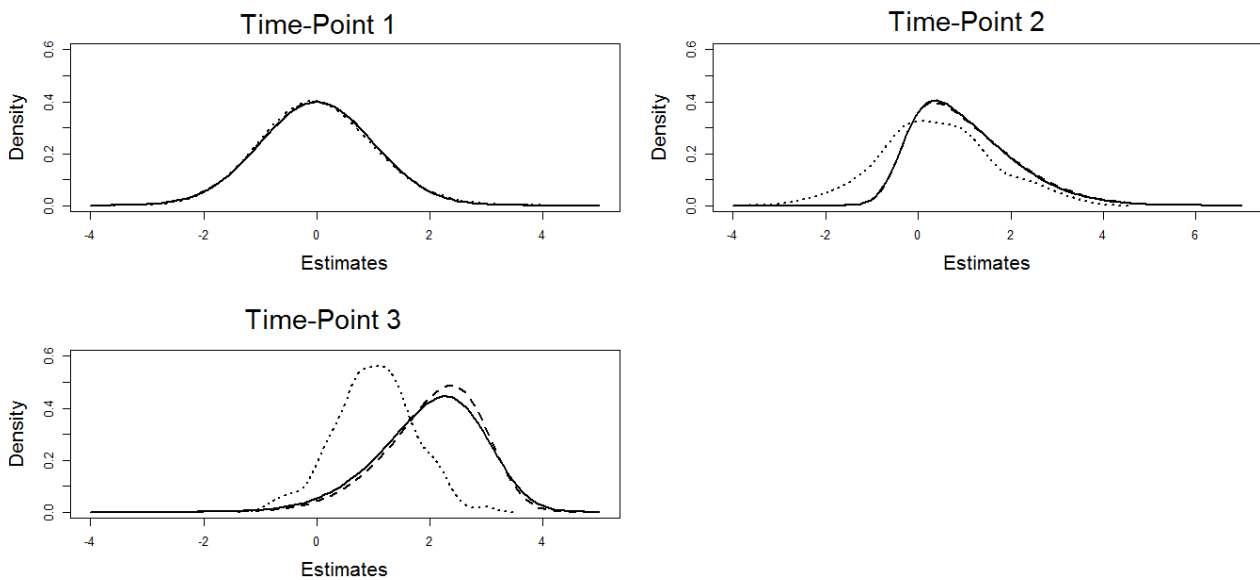


Figure 5.1: Estimated latent traits distributions. Legend: True distributions (solid line), Copula model (dashed line) and Cholesky model (dotted line)

Lastly, we compare the models by using the comparison statistics defined in the previous Chapters (see for example Chapter 4 equations (4.5.6)). The smaller is the value of statistic the better is the model fit. The copula model was chosen by all statistics.

5.5 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

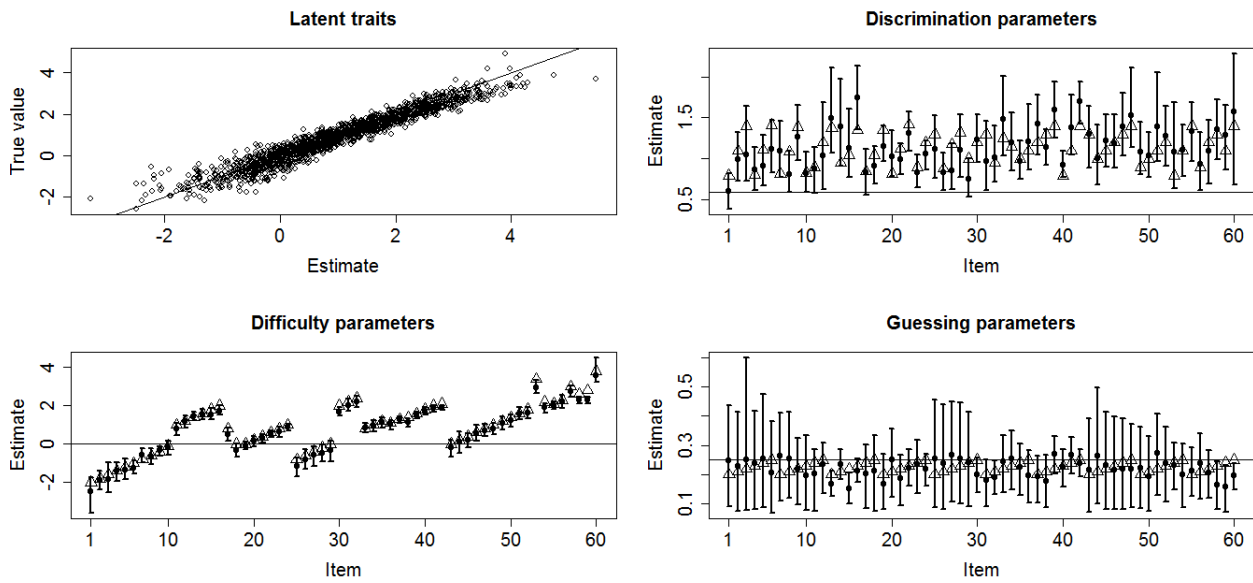


Figure 5.2: Estimates of latent traits and item parameters with 95% credibility intervals under the copula model. Legend: Circles denotes estimates, triangles denotes true values.

present analysis we consider a subset of the data analyzed in Chapters 2 and 3. 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 5.6 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 of the Chapter 3. 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 5.4 and 5.5 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 5.6 and 5.7 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 5.2.9 and the curves of the Cholesky model were simulated via model 5.2.6. 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 5.7. 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.

Figures 5.8 and 5.9 present the Bayesian p -value for the item parameters (see, for example, Chapter 4 section 4.5 for more details of this diagnostic) considering the two models. The results are similar and indicate that the two model fitted the items properly, with few exceptions. Figures 5.10 to 5.12 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.

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 5.3: Results for the estimated population parameters

	Copula				Cholesky		
	True value	Mean	SD	CI (95%)	Mean	SD	CI (95%)
μ_{θ_1}	.000	.000	–	–	.000	–	–
μ_{θ_2}	1.000	1.031	.070	[.889, 1.176]	.977	.068	[.841, 1.110]
μ_{θ_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]
ρ_{θ_1}	.800	.806	.011	[.783, .828]	.804	.015	[.776, .834]
ρ_{θ_2}	.700	.701	.031	[.640, .758]	.751	.025	[.699, .802]
γ_{θ_1}	.000	.028	.101	[-.160, .205]	.146	.136	[-.013, .457]
γ_{θ_2}	.800	.822	.091	[.606, .939]	.136	.084	[.012, .330]
γ_{θ_3}	-.500	-.594	.097	[-.745, -.396]	.078	.048	[.007, .210]

Table 5.4: 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 5.5: Models comparison: Simulation study

Model	DIC	EAIC	EBIC
Copula	37326.58	38049.87	44191.53
Cholesky	37998.86	39629.03	53471.32

Table 5.6: 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

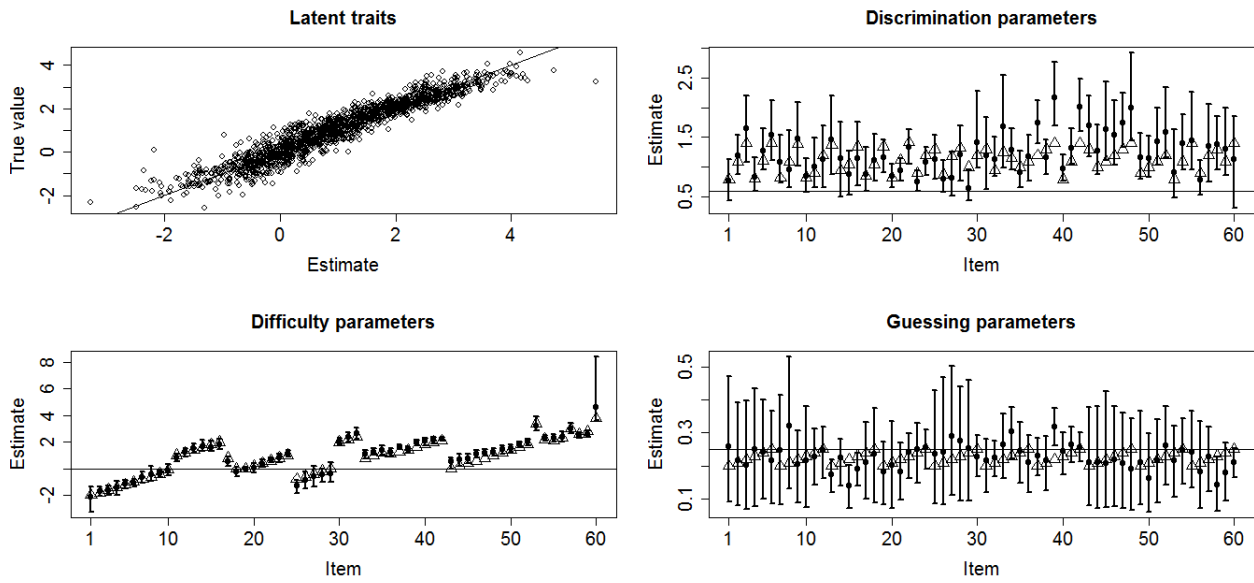


Figure 5.3: Estimates of latent traits and item parameters with 95% credibility intervals under the Cholesky model. Legend: Circles denotes estimates, triangles denotes true values.

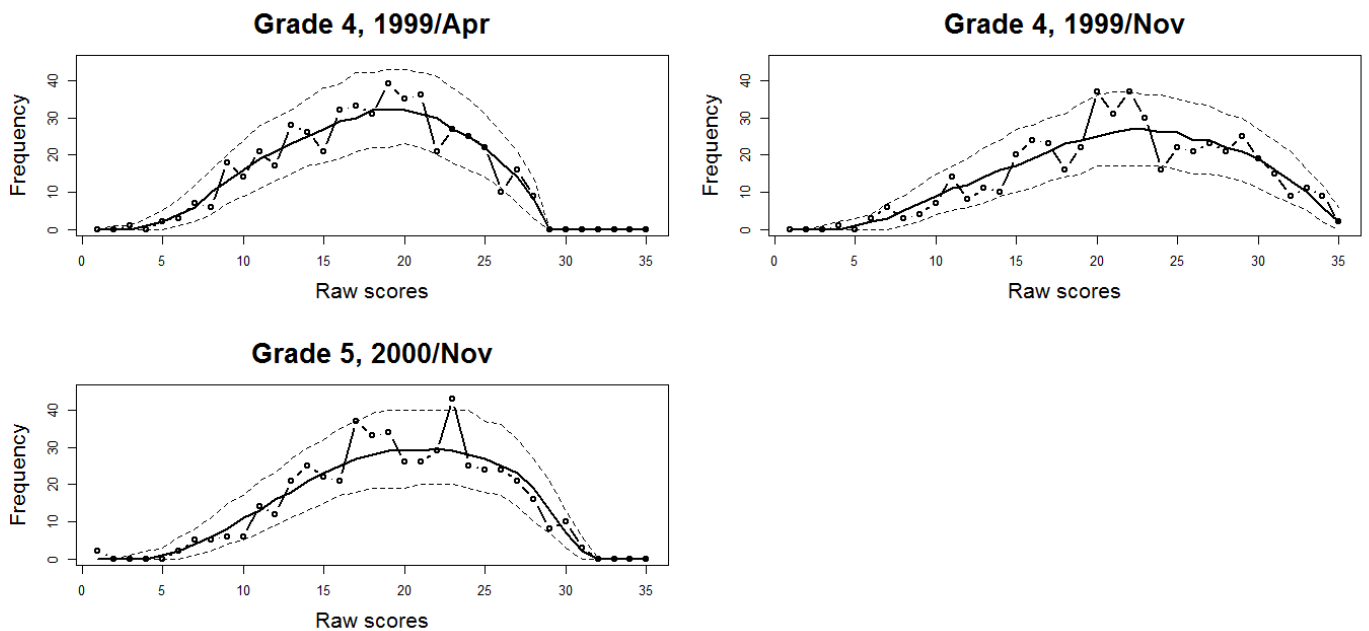


Figure 5.4: Observed and predicted scores distributions with 95% credibility intervals for the copula model

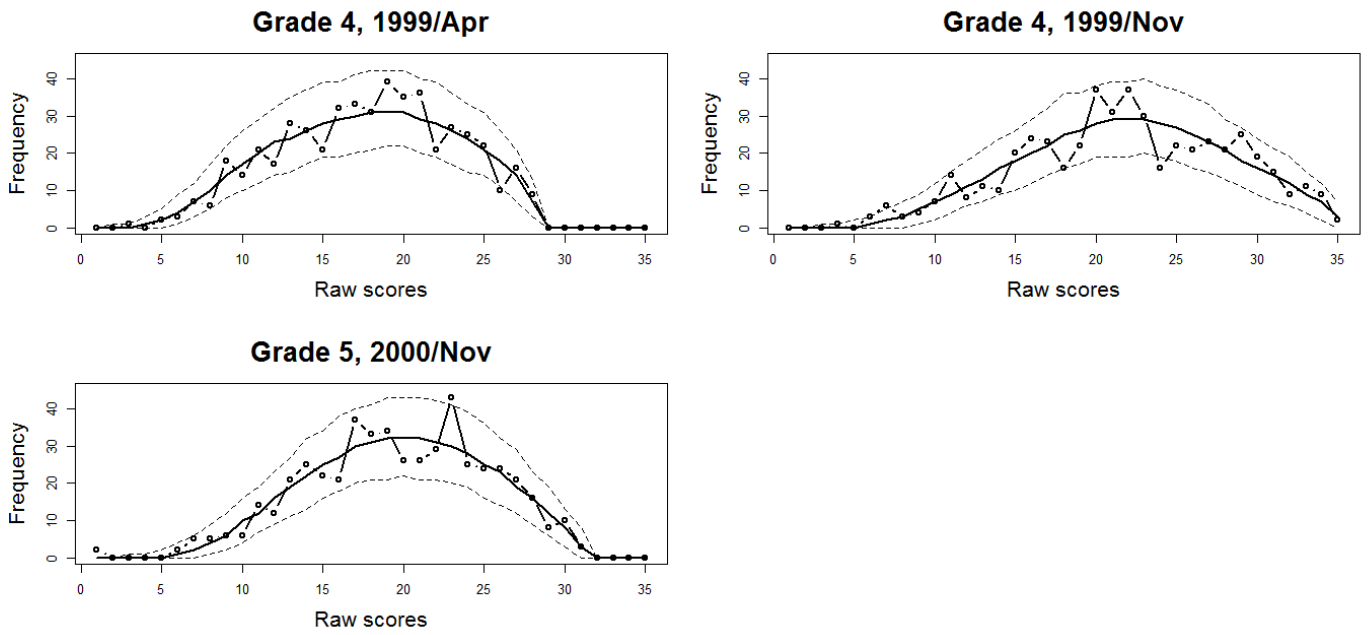


Figure 5.5: Observed and predicted scores distributions with 95% credibility intervals for Cholesky model

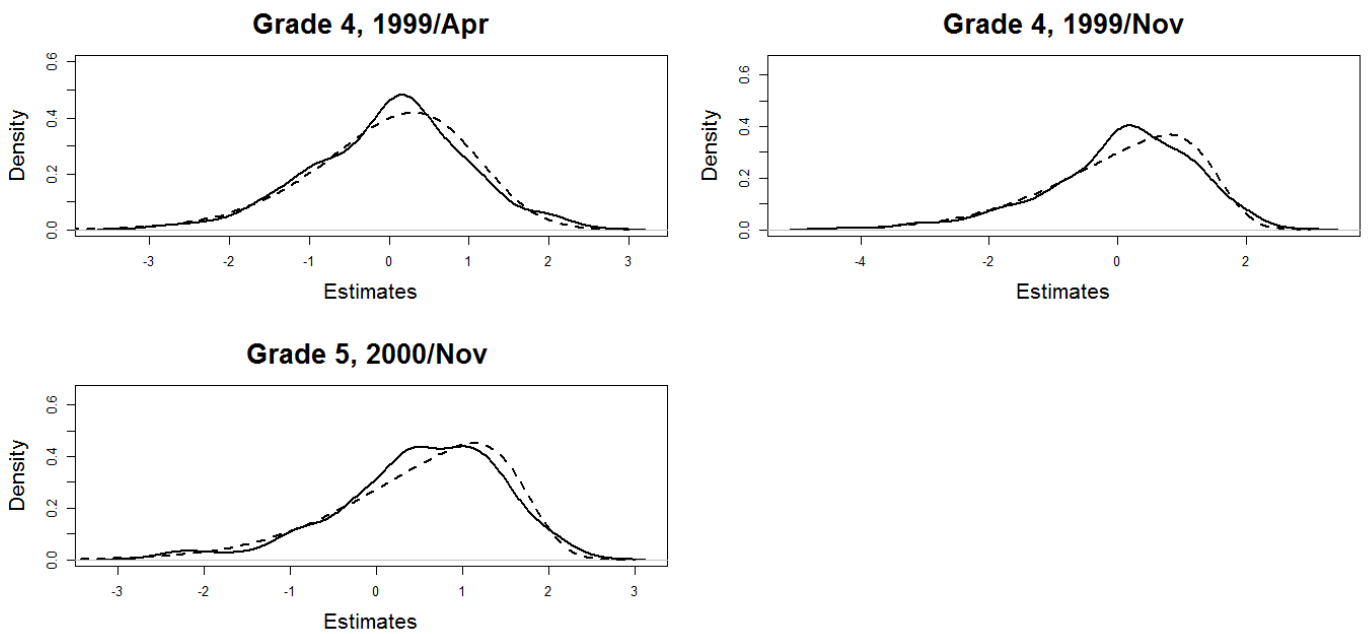


Figure 5.6: Estimated latent traits distributions with theoretical curves. Legend: Smoothed histogram (solid line), copula model (dashed line)

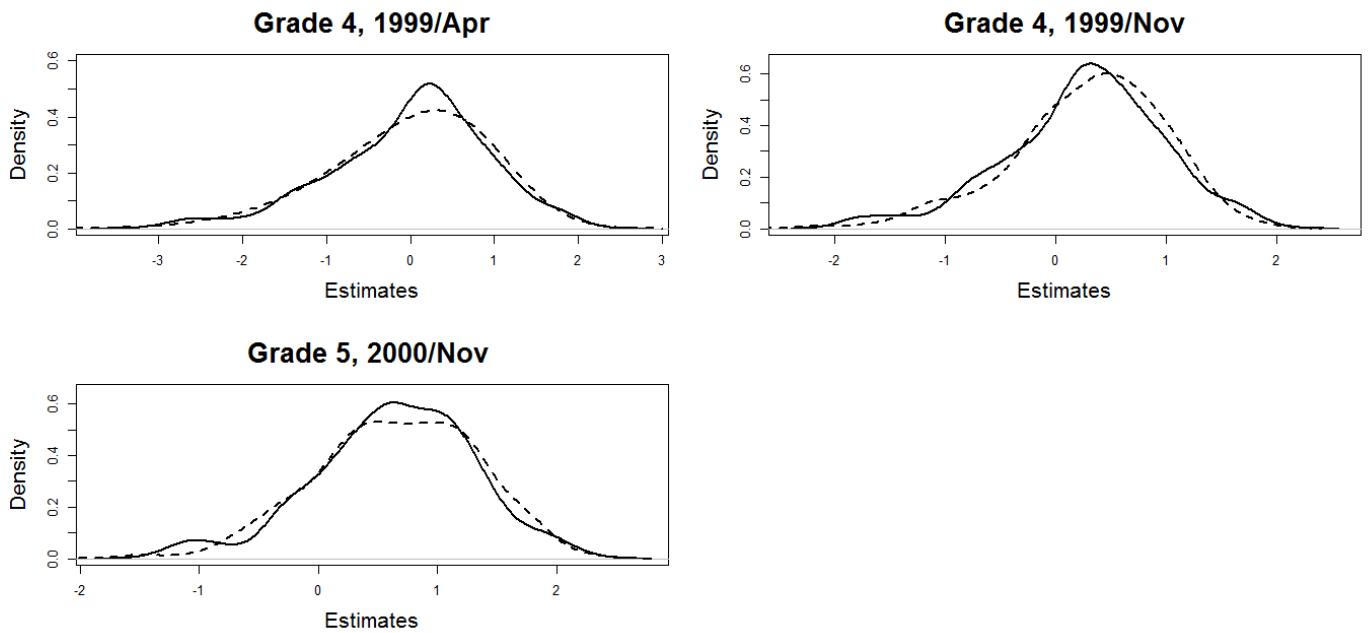


Figure 5.7: Estimated latent traits distributions with theoretical curves. Legend: Smoothed histogram (solid line), Cholesky model (dashed line)

Table 5.7: Estimates of the population parameters considering both copula and Cholesky models

	Copula			Cholesky		
	Mean	SD	CI (95%)	Mean	SD	CI (95%)
μ_{θ_1}	.000	—	—	.000	—	—
μ_{θ_2}	.053	.104	[-.124, .252]	.259	.059	[.133, .375]
μ_{θ_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]
ρ_{θ_1}	.810	.018	[.778, .853]	.847	.020	[.807, .882]
ρ_{θ_2}	.856	.029	[.797, .912]	.851	.024	[.800, .895]
γ_{θ_1}	-.500	.242	[-.858, .032]	-.526	.127	[-.743, -.229]
γ_{θ_2}	-.823	.128	[-.988, -.530]	-.366	.090	[-.519, -.158]
γ_{θ_3}	-.827	.113	[-.983, -.558]	-.319	.063	[-.433, -.188]

Table 5.8: Models comparison: Real data analysis

	DIC	EAIC	EBIC
Copula	54301.44	54633.80	65283.64
Cholesky	54267.23	54643.03	68541.70

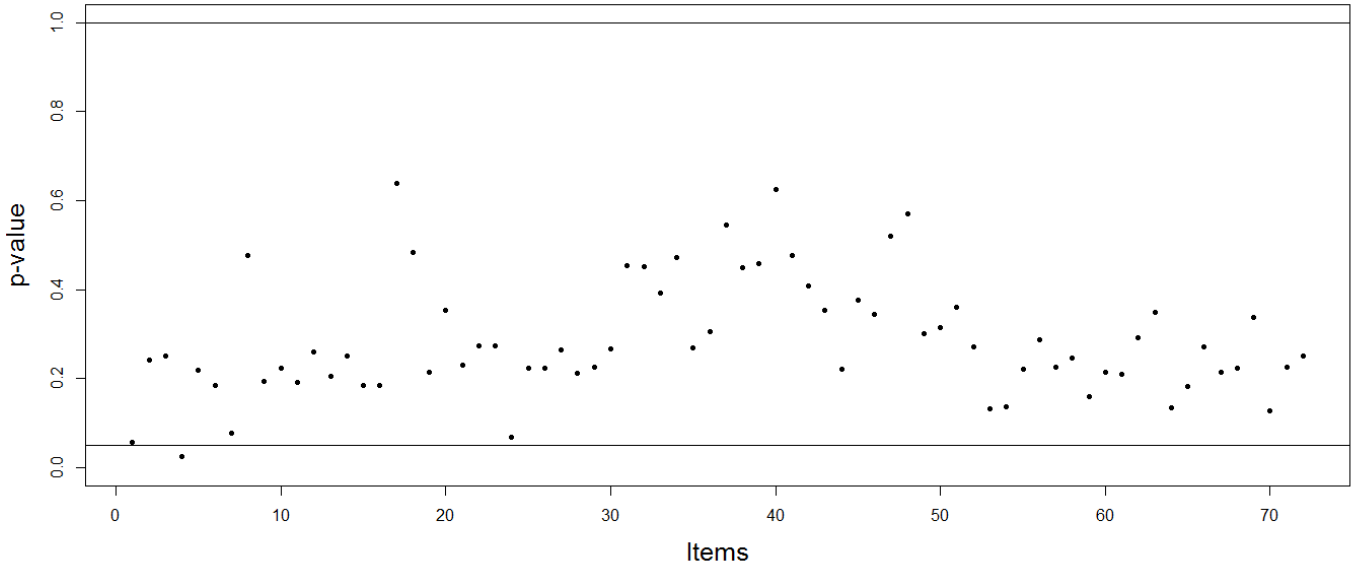


Figure 5.8: Bayesian p -values for item parameters considering the copula model

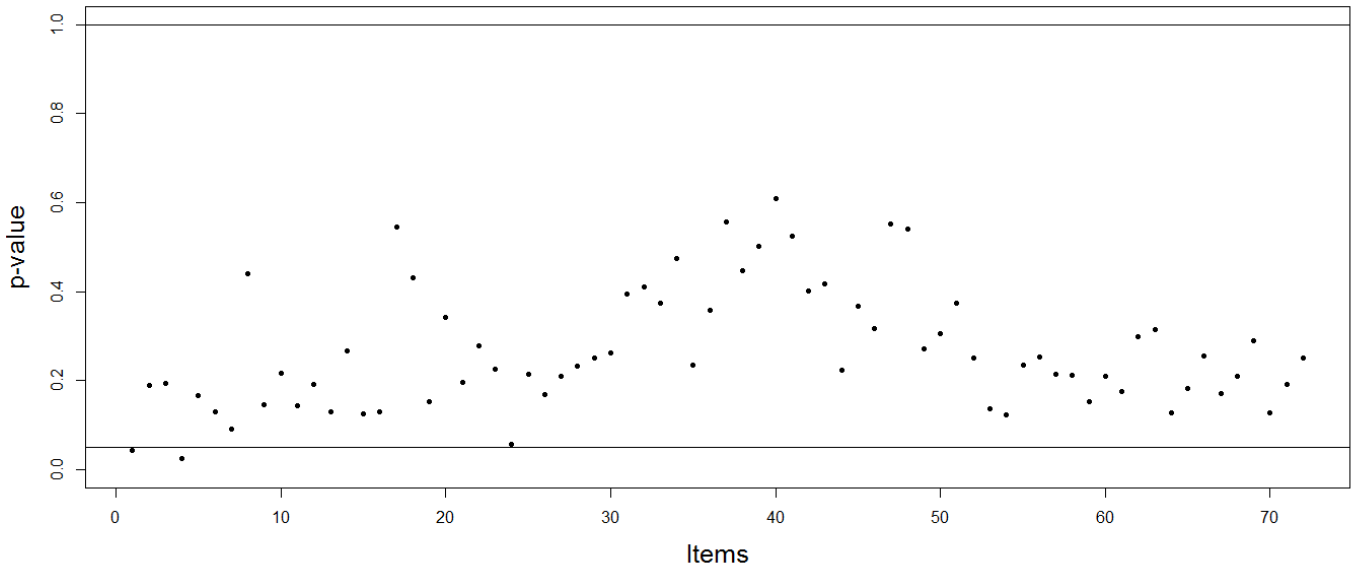


Figure 5.9: Bayesian p -values for item parameters considering the Cholesky model

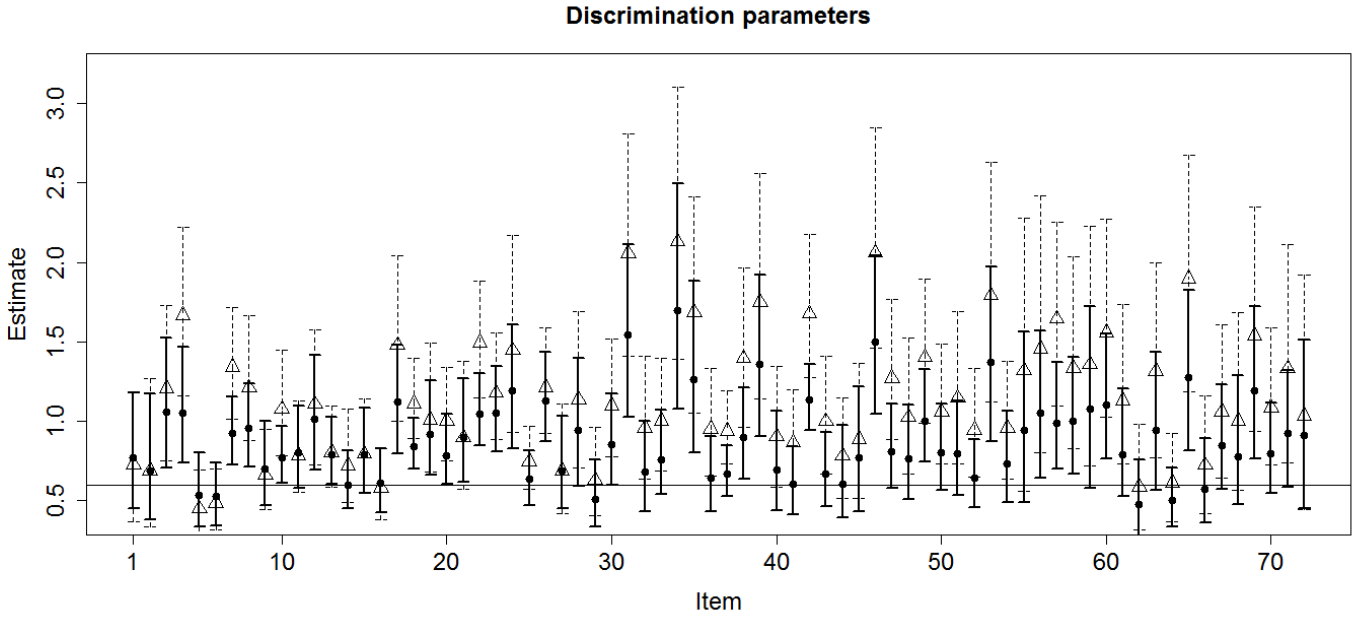


Figure 5.10: Discrimination parameters estimates.

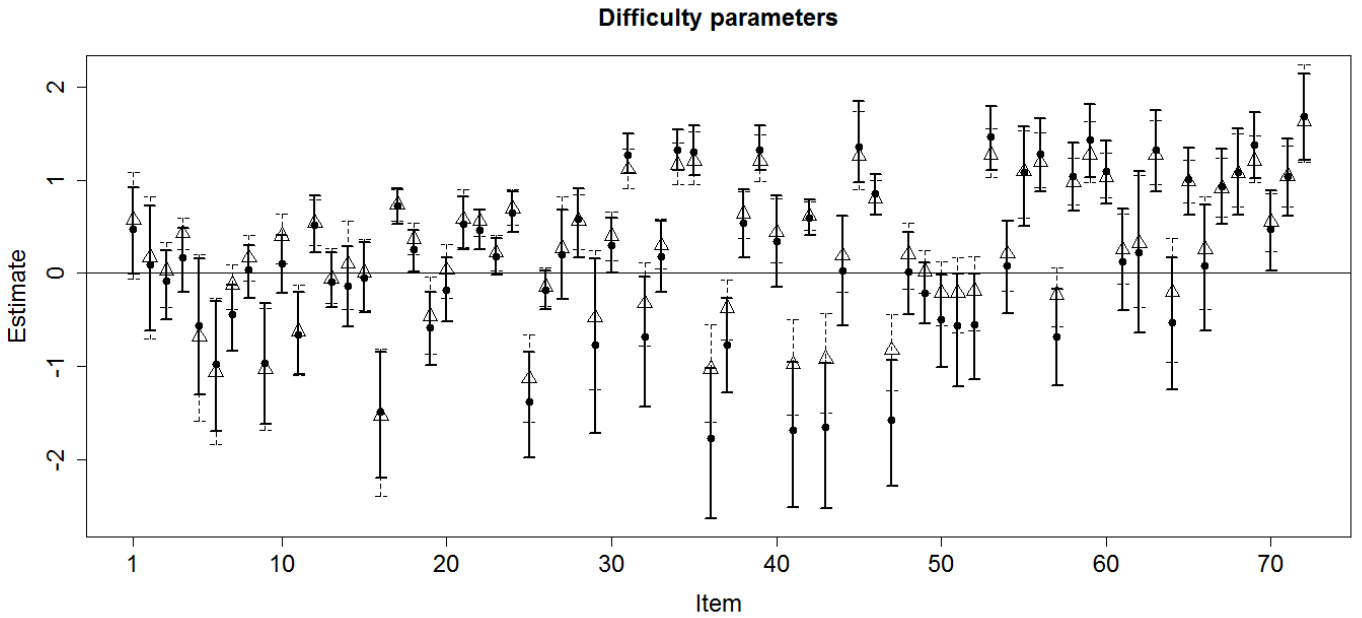


Figure 5.11: Difficulty parameters estimates.

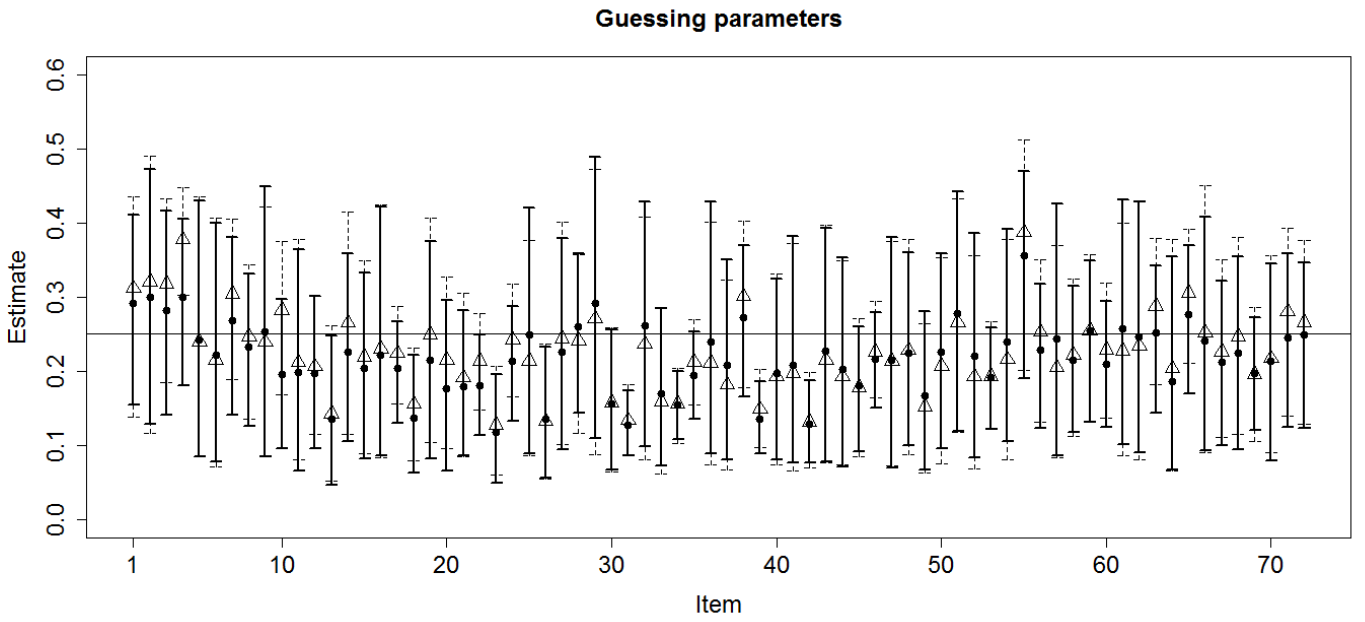


Figure 5.12: Guessing parameters estimates.

5.6 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, unlike to the approach developed in the previous chapters. 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.

Chapter 6

Final Conclusions

In the present thesis we proposed four classes of longitudinal IRT models. Such classes allows to consider skewed latent trait distributions, multiple group and unbalanced data structures and growth curves modeling. The approaches proposed here handle several situations of practical interest and they are promising alternatives for analyzing longitudinal IRT data. Specifically, our methodologies provide contributions in the following aspects:

- **Unbalanced data:** the antedependence modeling proposed in Chapter 2, 3 and 4, can handle easy dropouts and/or inclusions of subjects which are very common in longitudinal data studies;
- **Skewed latent trait distributions:** the assumption of skew-normality of the latent traits is more realistic than the common assumption of normality;
- **Covariance structure modeling:** both antedependence and copula approaches provide a very flexible modeling of structured covariance patterns. Our methodologies estimate only the necessary parameters (different from Azevedo et al. (2016) and Azevedo et al. (2015) in which more parameters are estimated) and can accommodate any covariance matrix;
- **Bayesian framework:** the antedependence modeling provides a conditional univariate structure for the latent traits. This allowed to develop a faster latent trait sampler, since no multivariate distributions are simulated in the MCMC algorithms. The developed FFBS sampler showed to efficient in terms of effective sample size compared to the Gibbs sampling. Furthermore, this algorithm can also be adapted to consider other latent trait distributions, through a suitable stochastic representation.

Therefore, our models are a promising alternatives to the usual ones in analyzing longitudinal IRT data with skewed latent trait distributions and multiple group structure. For future works, we intend to explore some extensions of the proposed models. We believe that the following aspects should be explored:

- **Another item response functions:** In tests with multiple choice alternatives or written items is more suitable to consider polytomous item response functions. These functions take into account the probability of response of each category (nominal or ordinal).
- **Another model classes:** The antedependence and/or copula modeling performance could be explored in other IRT model classes, such as, multidimensional, multivariate IRT models. These also require modeling of dependence structures related to the latent traits.

- **Regression structures:** Very often, in longitudinal studies, some subject-level background variables (as gender, age, education level of the parents, psychiatric characteristics) are collected. In this sense, regression structures for mean of the latent traits distribution, would allow to include explanatory information in the modeling to explain the latent traits variability, could be considered.
- **Another estimation methods:** The MCMC algorithms are very interesting option on the estimation of more complex IRT models under a Bayesian framework. However, its expensive computational cost, motivates the search for alternative methods. Frequentist methods, as the marginal maximum likelihood, see Bock and Aitkin (1981) could be explored. Another alternative it is to consider the so called conditional augmented data expectation maximization (CADEM) algorithm proposed by Azevedo and Andrade (2011). This is a variation of the EM algorithm and allows to fit complex models with reasonable facility and a low computational cost.

Bibliography

- Albert, J. (1992). Bayesian estimation of normal ogive item response curves using gibbs sampling. *Journal of educational statistics*, 17(3):251–269.
- Andrade, D. and Tavares, H. (2005). Item response theory for longitudinal data: population parameter estimation. *Journal of Multivariate Analysis*, 95(1):1–22.
- Andrade, D. F., Tavares, H. R., and Valle, R. C. (2000). *Item Response Theory: Concepts and Application (In portugues)*. 14^o SINAPE, ABE, first edition.
- 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.
- Azevedo, C. and Andrade, F. (2011). Cadem: A conditional augmented data em algorithm for fitting one parameter probit models. *Brazilian Journal of Probability and Statistics*, 27(2):245–262.
- 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.
- Azevedo, C., Bolfarine, H., and Andrade, D. (2012a). Parameter recovery for a skew-normal irt model under a bayesian approach: hierarchical framework, prior and kernel sensitivity and sample size. *Journal of Statistical Computation & Simulation*, 82(11):1679–1699.
- Azevedo, C. L., Andrade, D. F., and Fox, J.-P. (2012b). A bayesian generalized multiple group irt model with model-fit assessment tools. *Computational Statistics & Data Analysis*, 56(12):4399 – 4412.
- Azevedo, C. L., Fox, J.-P., and Andrade, D. F. (2015). Longitudinal multiple-group irt modelling: covariance pattern selection using mcmc and rjmc. *International Journal of Quantitative Research in Education*, 2(3-4):213–243.
- Azevedo, C. L. N. (2003). *Estimation Methods on Item Response Theory*. Master dissertation (In portugues), São Paulo, SP.
- 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.
- Azzalani, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 12:171–178.
- Baker, F. B. and Kim, S.-H. (2004). *Item response theory : Parameter Estimation Techniques*. Marcel Dekker, Inc, New York, NY, second edition.

- Bazán, J. L., Bolfarine, H., and Leandro, R. A. (2004). Bayesian estimation via mcmc for probit-normal model in item response theory. *Technical report, University of São Paulo*, (15):26.
- Bedford, T. and Cooke, R. (2002). Vines: a new graphical model for dependent random variables. *Annals of Statistics*, 30:1031–1068.
- 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.
- Bock, D. R. and Zimowski, M. F. (1997). *The Multiple Groups IRT*, in *Handbook of modern Item Response Theory*. Wim J. van der Linden and Ronald K. Hambleton eds, Springer-Verlag, New York, NY, first edition.
- 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.
- Braeken, J., Tuerlinckx, F., and Boeck, P. (2007). Copula functions for residual dependency. *Psychometrika*, 72(3):393–411.
- Branco, M. D. and Arellano-Valle, R. (2004). *Asymmetric Elipitical Distributions*. 16⁰ SINAPE, ABE, first edition.
- Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81:541–553.
- Cepeda-Cuervo, E. and Núñez-Antón, V. (2009). Bayesian modelling of the mean and covariance matrix in normal nonlinear models. *Journal of Statistical Computation and Simulation*, 79(6):837–853.
- Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*,. Wiley, Chichester, England, 1 edition.
- 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.
- Conaway, M. R. (1990). A random effects model for binary data. *Biometrics*, 46(2):317–328.
- Danaher, P. and Smith (2011). Modeling multivariate distributions using copulas: applications in marketing. *Marketing Science*, 30(1):4–21.
- Daniels, M. J. and Pourahmadi, M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, 89(3):553–566.
- Doebler, A. and Doebler, P. (2012). Building partially compensatory mirt models with copulae. working paper.
- 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.
- Frahm, G., Junker, M., and Szimayer, A. (2003). Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286.
- Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25.

- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis*, 15:183–202.
- Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo : Stochastic simulation for bayesian inference*. Chapman & Hall/CRC, London, second edition.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis*. Chapman & Hall, London, second edition.
- Genton, M. (2004). *Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Chapman & Hall, London, first edition.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice*. Chapman & Hall, London, first edition.
- Hambleton, R. K. and Swaminathan, H. (1985). *Item Response Theory : Principles and Applications*. Kluwer Academic Publishers, Boston, first edition.
- Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park : Sage publications, first edition.
- Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian journal of statistics*, 13:271–275.
- Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42(4):805–820.
- 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.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, 1 edition.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *ASME. J. Basic Eng*, 82(1):35–45.
- Kemper, H. C. G. and van 't Hof, M. A. (1978). Design of a multiple longitudinal study of growth and health in teenagers. *European Journal of Pediatrics*, 129(3):147–155.
- Kolen, M. J. and Brennan, R. L. (2004). *Teste Equating - Methods and Pratices*. Springer-verlarg, New York, second edition.
- Li, D. X. (1999). On default correlation: A copula function approach. *The RiskMetrics Group*.
- Lindsey, J. K. (1993). *Models for Repeated Measurements*. Oxford University Press, 2nd edition.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale : Lawrence Erlbaum Associates, Inc, first edition.
- Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA : Addison-Wesley, first edition.
- Marsman, M. (2014). *Plausible Values in Statistical Inference*. PhD Thesis University of Twente, Enschede., 1 edition.

- Montgomery, D. C. (2004). *Design and Analysis of Experiments*. Chapman & Hall, London, sixth edition.
- Murray, I., Ghahramani, Z., and MacKay, D. (2012). Mcmc for doubly-intractable distributions. ArXiv e-prints.
- Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer, 2 edition.
- Newton, H. (1988). *TIMESLAB: A time series analysis laboratory*. Pacific Grove, CA: Wadsworth & Brooks/Cole., 1 edition.
- Nojosa, R. T. (2001). *Multidimensional models on Item Response Theory*. Master dissertation (In portugues), Recife, PE, ufpe edition.
- Nunez-Anton, V. and Zimmerman, D. L. (2000). Modelinng nonstationary longitudinal data. *Biometrics*, 56(3):699–705.
- Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493.
- Pewsey, A. (2000). Problems of inference for azzalini’s skewnormal distribution. *Journal of Applied Statistics*, 27(7):859–870.
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690.
- Reckase, M. (2009). *Multidimensional Item Response Theory*. Springer New York, 1 edition.
- Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. *Markov Chain Monte Carlo in Praticce*, pages 45–57.
- Rodriguez, C. L. B. (2005). *Bayesian Inference on the Skew-normal Model*. Master dissertation (In portugues), São Paulo, SP, ime-usp edition.
- Sahu, S. K. (2002). Bayesian estimation and model choice in item response models. *Journal of Statistical Computation and Simulation*, 72(3):217–232.
- 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.
- Sinharay, S. (2006). A bayesian item fit analysis for unidimensional item response theory models. *British journal of mathematical and statistical psychology*, 59:429–449.
- Sinharay, S., Johnson, M. S., and Stern, H. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, 30(4):298–321.
- 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.
- 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.

- 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.
- Spielberger, C. (1989). *State-Trait Anxiety Inventory: Bibliography*. Palo Alto, CA: Consulting Psychologists Press, 2 edition.
- 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.
- Tavares, H. R. (2001). *Item Response Theory for Longitudinal Data*. PhD thesis (In portugues), São Paulo, SP, ime-usp edition.
- Tavares, H. R. and Andrade, D. F. (2006a). Item response theory for longitudinal data: Item and population ability parameters estimation. *TEST*, 15(1):97–123.
- Tavares, H. R. and Andrade, D. F. (2006b). Latent distribution parameter estimation and approximation. *personal communication*.
- Xue-Kun Song, P. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320.
- Zimmerman, D. and Núñez-Antón, V. (2009). *Antedependence Models for Longitudinal Data*. Chapman and Hall/CRC, 1 edition.

Appendix A

Further results on SVE algorithms

A.1 SVE algorithm for Heteroscedastic Toeplitz matrix

Algorithm A.1 present an example of a multidimensional SVE. For illustration, we consider symmetric model with HT matrix.

Algorithm A.1 The SVE algorithm with oversampling to sample a correlation parameter considering HT matrix

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

Require: A function `HT.matrix()` to build the HT matrix

1: **for** $t = 1$ to $T - 1$ **do**

2: **for** $g = 1$ to G **do**

3: Draw $\rho_{\theta_{gt}}^{(m)} \sim p(\rho_{\theta_t})$

4: Draw $\boldsymbol{\theta}_{.ts}^{(m)}$ from the model (2.2.8)

5: Compute $r_t(\boldsymbol{\theta}_{.ts}^{(m)})$ the sample correlation of order t

6: **end for**

7: Choose the $\rho_{\theta_{ts}}^{(m)}$ and $\boldsymbol{\theta}_{.ts}^{(m)}$ whose $r_t(\boldsymbol{\theta}_{.ts}^{(m)})$ is closest to $r_t(\boldsymbol{\theta}_{.t}^{(m-1)})$

8: Set $\rho_{\theta_t}^{(m)} = \rho_{\theta_{ts}}^{(m)}$ and $\boldsymbol{\theta}_{.t}^{(m)} = \boldsymbol{\theta}_{.ts}^{(m)}$ the candidate values

9: **end for**

10: Set $\boldsymbol{\rho}_\theta^{(m)} = (\rho_{\theta_1}^{(m)}, \dots, \rho_{\theta_{T-1}}^{(m)})$

11: Build the HT proposed matrix $\Sigma_{\boldsymbol{\rho}_\theta}^{(m)}$ using `HT.matrix()`

12: Perform the Cholesky decomposition of $\Sigma_{\boldsymbol{\rho}_\theta}^{(m)}$ to obtain the matrices $\mathbf{L}^{(m)}$ and $\mathbf{D}^{(m)}$

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

14: **if**

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

then

15: $\boldsymbol{\rho}_\theta^{(m-1)} = \boldsymbol{\rho}_\theta^{(m)}$

16: **end if**

A.2 Some useful R functions

Some functions R functions used to implement the estimation algorithms are presented below.

A.2.1 Cholesky decomposition

This function performs the Cholesky decomposition of positive definite matrix **A** returning a lower triangular matrix **L** and diagonal matrix **D**.

```
chol.LDL<-function(A){
ch<-chol(A)
dd<-diag(ch)
T<-t(ch/dd)

L_mat<-solve(T)
D_mat<-diag(dd^2)
return(list(T=T,D_mat=D_mat,L_mat=L_mat))
}
```

A.2.2 Functions to build structured matrices

First-order Heteroscedastic Autoregressive

```
ARH1.matrix<-function(sigma2,rho){
m<-length(sigma2)
Sigma<-gama.arma(toeplitz(0:(m-1)),1,rho,0)
R<-(1/Sigma[1,1])*Sigma
mat_aux1<-sqrt(matrix(sigma2[matrix(rep(1:m,m),m,m,byrow=T)],m,m))
mat_aux2<-sqrt(matrix(sigma2[matrix(rep(1:m,m),m,m,byrow=F)],m,m))
Sigma<-R*mat_aux1*mat_aux2
diag(Sigma)<-sigma2
return(Sigma)
}
```

First-order Heteroscedastic Autoregressive Moving-Average

```
gama.arma<-function(k,sigma2,rho1,rho2){
gama.0<- sigma2*((1 + 2*rho2*rho1 + rho2^2) / (1-rho1^2))
gama.1<-sigma2*((1+rho2*rho1)*(rho1+rho2)/(1-rho1^2))
gama.k<-(rho1^(k-1))*gama.1
return(gama.0*(k==0) + gama.1*(k==1) + gama.k*(k>=2))
# returns the autocovariance function of an ARMA process
}
```

```
ARMAH11.matrix<-function(sigma2,rho1,rho2){
vT<-length(sigma2)
R<-rho2*cov2cor(gama.arma(toeplitz(c(0,0:(vT-2))),1,rho1,0))
mat_aux1<-sqrt(matrix(sigma2[matrix(rep(1:vT,vT),vT,vT,byrow=T)],vT,vT))
```

```

Sigma<-R*mat_aux1*t(mat_aux1)
diag(Sigma)<-sigma2
return(Sigma)
}

```

Heteroscedastic Toeplitz

```

toeplitz.matrix<-function(sigma2,rho){
m<-length(sigma2)
R<-toeplitz(c(1,rho))
mat_aux1<-sqrt(matrix(sigma2[matrix(rep(1:m,m),m,m,byrow=T)],m,m))
mat_aux2<-sqrt(matrix(sigma2[matrix(rep(1:m,m),m,m,byrow=F)],m,m))
return(R*mat_aux1*mat_aux2)
}

```

Antedependence Matrix

```

Prodacf<-function(rho,h){
acf<-c()
acf[1]<-1
for(i in 2:h){
acf[i]<-rho[i-1]*acf[i-1]
}
return(acf) # returns the autocorrelation function of an AD process
}

```

```

Prod.matrix<-function(sigma2,rho){
m<-length(sigma2)
R<-toeplitz(Prodacf(rho,m))
mat_aux1<-sqrt(matrix(sigma2[matrix(rep(1:m,m),m,m,byrow=T)],m,m))
mat_aux2<-sqrt(matrix(sigma2[matrix(rep(1:m,m),m,m,byrow=F)],m,m))
Sigma<-R*mat_aux1*mat_aux2
return(Sigma)
}

```

Function `toeplitz()` is a standard R function.

Appendix B

Item parameters of the simulation studies

Table B.1: Item parameters of the simulation studies of the Chapters 2 and 3

Item	Test	Parameters			Item	Test	Parameters			Item	Test	Parameters		
		a_i	b_i	c_i			a_i	b_i	c_i			a_i	b_i	c_i
1	1-2	.70	-1.50	.20	41	3-4	.98	1.76	.24	81	5-6	.78	2.03	.22
2	1-2	.73	-1.30	.21	42	3-4	1.20	1.98	.25	82	5-6	1.43	2.21	.23
3	1-2	.80	-.90	.22	43	3-4	1.30	.16	.20	83	5-6	1.73	.72	.24
4	1-2	.75	1.00	.23	44	3-4	.86	.64	.21	84	5-6	1.10	1.11	.25
5	1-2	.70	-1.95	.24	45	3-4	.82	.98	.22	85	5-6	.70	1.39	.20
6	1-2	.90	1.60	.25	46	3-4	.88	1.56	.23	86	5-6	.80	1.87	.21
7	1-2	.93	1.54	.20	47	3-4	.90	1.76	.24	87	5-6	.80	2.03	.22
8	1-2	.70	.97	.21	48	3-4	1.34	2.00	.25	88	5-6	1.85	2.23	.23
9	1-2	.92	1.98	.22	49	3-4	1.20	2.56	.20	89	5-6	1.43	2.69	.24
10	1-2	.77	1.55	.23	50	3-4	1.10	2.64	.21	90	5-6	1.13	2.75	.25
11	1-2	1.00	.65	.24	51	3-4	.84	2.34	.22	91	5-6	.90	2.51	.20
12	1-2	.74	.10	.25	52	3-4	.90	2.40	.23	92	5-6	.80	2.56	.21
13	1-2	.80	.14	.20	53	3-4	1.30	2.33	.24	93	5-6	1.73	2.50	.22
14	1-2	.83	1.10	.21	54	3-4	.83	1.78	.25	94	5-6	.85	2.05	.23
15	1-2	.90	-1.09	.22	55	3-4	1.57	1.55	.20	95	5-6	2.53	1.86	.24
16	1-2	.78	.65	.23	56	3-4	.92	2.40	.21	96	5-6	.80	2.56	.25
17	1-2	.88	.55	.24	57	3-4	.78	.96	.22	97	5-6	.80	1.37	.20
18	1-2	.90	1.30	.25	58	3-4	.80	.33	.23	98	5-6	.90	.86	.21
19	1-2	.95	1.45	.20	59	3-4	.99	1.20	.24	99	5-6	.81	1.57	.22
20	1-2	.85	1.45	.21	60	3-4	.94	1.87	.25	100	5-6	.98	2.12	.23
21	2-3	.86	1.75	.22	61	4-5	.93	1.62	.20	101	6	.87	1.92	.24
22	2-3	.72	-.60	.23	62	4-5	.86	.50	.21	102	6	.88	1.00	.25
23	2-3	.77	-.98	.24	63	4-5	.88	.92	.22	103	6	.89	1.34	.20
24	2-3	.80	1.00	.25	64	4-5	1.30	1.45	.23	104	6	1.73	1.78	.21
25	2-3	1.20	2.34	.20	65	4-5	.92	1.76	.24	105	6	1.10	2.03	.22
26	2-3	1.00	2.00	.21	66	4-5	1.20	2.00	.25	106	6	1.43	2.23	.23
27	2-3	.95	1.97	.22	67	4-5	1.30	2.98	.20	107	6	1.73	3.03	.24
28	2-3	.97	2.70	.23	68	4-5	1.34	3.10	.21	108	6	1.85	3.13	.25
29	2-3	1.32	2.10	.24	69	4-5	1.38	3.40	.22	109	6	1.97	3.38	.20
30	2-3	1.23	2.36	.25	70	4-5	1.10	2.30	.23	110	6	1.13	2.47	.21
31	2-3	.98	1.67	.20	71	4-5	1.40	3.54	.24	111	6	2.02	3.49	.22
32	2-3	1.32	2.30	.21	72	4-5	1.10	2.65	.25	112	6	1.13	2.76	.23
33	2-3	.73	-.50	.22	73	4-5	1.20	2.80	.20	113	6	1.43	2.89	.24
34	2-3	.90	2.20	.23	74	4-5	1.60	2.54	.21	114	6	2.62	2.67	.25
35	2-3	.95	2.76	.24	75	4-5	2.50	3.50	.22	115	6	3.29	3.46	.20
36	2-3	.89	1.89	.25	76	4-5	1.37	3.30	.23	116	6	1.94	3.30	.21
37	2-3	1.50	2.32	.20	77	4-5	1.20	3.00	.24	117	6	1.43	3.05	.22
38	2-3	1.45	2.00	.21	78	4-5	1.10	2.97	.25	118	6	1.13	3.03	.23
39	2-3	1.10	2.65	.22	79	4-5	1.20	1.98	.20	119	6	1.43	2.21	.24
40	2-3	1.56	1.00	.23	80	4-5	1.40	4.00	.21	120	6	2.02	3.87	.25

Table B.2: Item parameters of the simulation study of the Chapter 5

Item	Test	Parameters			Item	Test	Parameters		
		a_i	b_i	c_i			a_i	b_i	c_i
1	1	.80	-2.00	.20	31	2	1.30	2.20	.20
2	1	1.10	-1.80	.21	32	2	.95	2.40	.21
3	1	1.40	-1.60	.22	33	2	1.25	.80	.22
4	1	.81	-1.40	.23	34	2	1.15	1.00	.23
5	1	1.11	-1.20	.24	35	2	1.00	1.20	.24
6	1	1.41	-1.00	.25	36	2	1.10	1.10	.25
7	1	.82	-.80	.20	37	2-3	1.20	1.30	.20
8	1	1.09	-.60	.21	38	2-3	1.30	1.40	.21
9	1	1.39	-.40	.22	39	2-3	1.40	1.60	.22
10	1	.83	-.10	.23	40	2-3	.80	1.80	.23
11	1	.90	1.00	.24	41	2-3	1.10	2.00	.24
12	1	1.20	1.20	.25	42	2-3	1.40	2.10	.25
13	1	1.38	1.40	.20	43	3	1.30	.00	.20
14	1	.95	1.60	.21	44	3	1.00	.20	.21
15	1	1.05	1.80	.22	45	3	1.10	.40	.22
16	1	1.35	2.00	.23	46	3	1.20	.60	.23
17	1	.85	.80	.24	47	3	1.30	.80	.24
18	1	1.05	.00	.25	48	3	1.40	1.00	.25
19	1-2	1.35	.00	.20	49	3	.90	1.20	.20
20	1-2	.82	.20	.21	50	3	1.00	1.40	.21
21	1-2	1.12	.40	.22	51	3	1.10	1.60	.22
22	1-2	1.42	.60	.23	52	3	1.20	1.80	.23
23	1-2	.90	.80	.24	53	3	.80	3.40	.24
24	1-2	1.20	1.00	.25	54	3	1.10	2.20	.25
25	2	1.30	-.80	.20	55	3	1.40	2.10	.20
26	2	.88	-.60	.21	56	3	.90	2.30	.21
27	2	1.18	-.40	.22	57	3	1.20	3.00	.22
28	2	1.32	-.20	.23	58	3	1.30	2.60	.23
29	2	1.00	.00	.24	59	3	1.10	2.80	.24
30	2	1.20	2.00	.25	60	3	1.40	3.80	.25