UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

ALEXANDRE HENRIQUE CARVALHO MARQUES

MULTIPLE FACTOR ANALYSIS MODEL WITH SCALE MIXTURE OF NORMAL DISTRIBUTIONS IN THE LATENT FACTORS

Recife
2018

ALEXANDRE HENRIQUE CARVALHO MARQUES

MULTIPLE FACTOR ANALYSIS MODEL WITH SCALE MIXTURE OF NORMAL DISTRIBUTIONS IN THE LATENT FACTORS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Área de concentração: Matemática-estatística.

Orientador: Prof. Dr. Aldo William Medina Garay.

Coorientador: Prof. Dr. Francisco José de Azevêdo Cysneiros.

Recife

2018

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

M357m Marques, Alexandre Henrique Carvalho
Multiple factor analysis model with scale mixture of normal distributions in the latent factors / Alexandre Henrique Carvalho Marques. – 2018.
100 f.: il., fig., tab.

Orientador: Aldo William Medina Garay.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2018.
Inclui referências e apêndices.

1. Estatística. 2. Análise fatorial. I. Garay, Aldo William Medina (orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2018-120

ALEXANDRE HENRIQUE CARVALHO MARQUES

MULTIPLE FACTOR ANALYSIS MODEL WITH SCALE MIXTURE OF NORMAL DISTRIBUTIONS IN THE LATENT FACTORS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 27/07/2018

BANCA EXAMINADORA

Prof. Dr. Aldo William Medina Garay (Orientador)
Universidade Federal de Pernambuco

Prof. Dr. Getúlio José Amorim do Amaral (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Victor Hugo Lachos Dávila (Examinador Externo)
University of Connecticut - USA

This research is dedicated to my amazing uncle Otávio and aunt Lena and to my awesome cousins Gabriel, Lúcio and Mila.

Acknowledgments

This dissertation had many contributors that were essential for its development. Firstly, I should mention the names of my adviser Prof. Dr. Aldo William Medina Garay and my co-adviser Prof. Dr. Francisco José de Azevêdo Cysneiros. I should thank them for all the effort and patience they had during the orientations. I should also thank all other professors of the Department of Statistics of the UFPE. They all contributed significantly for my advances in statistics in the last two years. Also, I would like to thank the CAPES and CNPq for the financial support.

I would like to thank my family for all the support they gave me throughout the entire time I was enrolled in the master course in statistics. My uncles and aunts were fundamental for my persistence during this period. Finally, I thank my father, mother and sister for their dedication and indispensable help.

Abstract

Statistical tools for modeling covariance structures have been shown useful in Medicine for studies in genetics. In that context, factor analysis models stand out for its ability in identifying latent factors capable of reducing data dimensionality and explaining observed variability. Usually, latent factors are interpreted as unobserved physiological mechanisms underlying the studied phenomenon. Confirmatory factor analysis models are characterized by allowing the researcher to pre-specify model's elements, as for example, the number of latent factors, the loading matrix structure and linear restrictions on the parameters. Those models allow the validation of hypothesis in gene co-expression studies. Confirmatory factor analysis models under normality assumption for the data are well consolidated in the literature. Our aim is to develop a more general class capable of integrate several independent populations extending the data's normality assumption to a more flexible class of distributions, the class of scale mixture of normal (SMN). The class of scale mixture of normal includes, as special cases, the normal distribution and distributions with heavy tails as the t-Student, contaminated normal ans slash. This model allows to specify parameter restrictions, which leads to important particular cases of covariance structures, making it more flexible in its specification and distributional assumptions. Model identifiability is studied, with necessary and/or sufficient conditions for parameter identification being presented. To estimate the model's parameters we propose an ECM algorithm and the estimators' performance in finite samples is evaluated through Monte Carlo simulation studies. We conclude the study with an illustration considering a confirmatory model for the pathological dynamic of pancreas cancer based on actual gene expression data.

Keywords: Multiple Confirmatory Factor Analysis. Identifiability. ECM algorithm. Class of scale mixture of normal distributions (SMN).

Resumo

Ferramentas estatísticas voltadas para a modelagem de estruturas de covariâncias têm se mostrado úteis em medicina para estudos genéticos. Nesse contexto, modelos de análise fatorial destacam-se por sua habilidade em identificar fatores latentes capazes de reduzir a dimensionalidade dos dados e explicar a variabilidade observada. Comumente, fatores latentes são interpretados como mecanismos fisiológicos não observáveis subjacentes ao fenômeno estudado. Modelos de análise fatorial confirmatória caracterizam-se por possibilitar ao pesquisador a pré-especificação de elementos do modelo, como por exemplo, o número de fatores latentes, a estrutura da matriz de *loadings* e restrições lineares nos parâmetros. Tais modelos permitem a validação de hipotéses em estudos de coexpressão gênica. Modelos de análise fatorial confirmatório sob suposição de normalidade de dados estão bem consolidados na literatura. Nosso objetivo é desenvolver uma classe mais geral capaz de integrar várias populações independentes estendendo a suposição de normalidade de dados para uma classe mais flexível de distribuições, a classe de misturas de escala da distribuição normal (SMN). A classe SMN contém, como casos especiais, a distribuição normal e distribuições com caudas pesadas tais como t-*Student*, normal contaminada e *slash*. Este modelo permite especificar restrições nos parâmetros, as quais levam a importantes casos particulares de estruturas de covariância, tornando-o mais flexível em sua especificação e em suas suposições distribucionais. A identificabilidade do modelo é estudada e condições necessárias e/ou suficientes para identificação dos parâmetros são apresentadas. Para a estimação dos parâmetros do modelo propomos um algoritmo ECM e a performance dos estimadores em amostras finitas é avaliada através de estudos de simulação de Monte Carlo. Finalizamos nosso estudo com uma ilustração considerando o modelo confirmatório para a dinâmica patológica do câncer de pâncreas utilizando dados reais de expressão gênica.

Palavras chave: Análise Fatorial Confirmatória Múltipla. Identificabilidade. Algoritmo ECM. Classe de misturas de escala da distribução normal (SMN).

List of figures

List of tables

Content

1 Introduction

1.1 Resumo da seção

Nesta seção introdutória nós apresentamos a motivação do problema de pesquisa e elementos essenciais de probabilidade e estatística para o entendimento das técnicas desenvolvidas nas seções subsequêntes. A motivação surge da necessidade de testar hipóteses de causalidade no campo da medicina, especialmente no campo da genética, onde avanços recentes em tecnologia laboratorial tem possibilitado o acumulo de grandes volumes de dados. Neste ponto, a Análise fatorial (AF) apresentasse como um técnica multivariada relevante para modelagem em genética, como tem se observado na literatura recente. Neste sentido, nossas contribuições surgem a partir de um novo modelo de AF que estende a suposição de normalidade dos dados para a classe de misturas de escala de normal (SMN). Na seção final, nós introduzimos a classe SMN, a qual será a base para nosso novo modelo de AF.

1.2 Motivation

Modern advances in laboratory technology for processing genetic material have driven the medical sciences towards the use of new methodologies for planning experiments (KERR, 2001; GERMAIN *et al.*, 2011). Two important modern technologies are RNA sequencing (RNA-seq) and microarrays, which allow the simultaneous measurement of thousands of biological parameters in cell populations, with the possibility of covering the entire transcriptome, *i.e.* all RNA molecules in the cell (KERR, 2001). The enlarged amount of data generated in modern medical researches contributes also for the generation of new theories in molecular cell biology, genetics and immunology (KERR, 2001; RIECKMANN *et al.*, 2017).

Models for the immune system have been conceived with the aid of mathematics through differential equations (PERELSON, 1989), bio-informatics using neural network (HOFFMANN, 1986) and methods based on simulation (GERMAIN *et al.*, 2011), and statistics with latent variable models (ROY *et al.*, 2014; BROWN *et al.*, 2015; BUETTNER *et al.*, 2017; DE VITO, 2016; WANG and PARMIGIANI, 2018). Focusing on statistical methods, network models for the immune system are commonly based on data gathered in experiments measuring gene expression, mainly with outputs of RNA-seq and microarray essays (BROWN *et al.*, 2015). Those models are called gene co-expression models (ROY *et al.*, 2014).

Usually, gene co-expression models are conceived in terms of latent factors, which are thought of as unobserved biological pathways [1] (DE VITO, 2016). Biological pathways can only be directly observed in laboratory experiments, which are over-simplified versions of physiological processes and usually leads to reductionist conclusions about the immune system (GERMAIN *et al.*, 2011). The statistical formulation of biological pathways as latent random variables allows for a thorough and more realistic analysis since it uses data directly measured on the actual system being modeled, the human physiology. (DE VITO, 2016).

Factor analysis models (FA) are an important class of latent variable models commonly applied for the exploration of biological pathways using microarray data (BROWN *et al.*, 2015; BUETTNER *et al.*, 2017). In the context of gene co-expression modeling, refinements of the FA model were undertaking by, for example, Brown *et al.* (2015) and Buettner *et al.* (2017) with the aim of segregate random noise due to batch effects from biological signal in order to effectively infer new biological pathways and improve gene set annotation, *i.e.* to refine the knowledge of genes' biological function.

Multiple factor analysis (MFA) models (JÖRESKOG, 1971), which are extensions of FA models oriented to the simultaneous analysis of several independent data sets, has also been shown an important statistical tool for modeling gene co-expression with microarray data from different tissues or from independent experiments in meta-analysis studies [2] (DE VITO, 2016; WANG and PARMIGIANI, 2018). De Vito (2016) explored a particular MFA model applied to the refinement of biological signals using microarray data in meta-analysis studies. Wang and Parmigiani (2018) studied a MFA model towards meta-analysis studies producing gene expression data through different methods. The authors discussed how to combine the different source of data in order to generate reliable scores for genes.

Despite the importance of statistical methods designed to explore and to reveal new pathways in human physiology, there is a lack of confirmatory models for testing pre-specified theories arising in medical researches. For that aim, the most commonly

[1] A biological pathway is a cascade of chemical and physical events connecting molecules and cells in complex networks for the control of physiological functions.

[2] The term meta-analysis refers to the simultaneous analysis of several independent data sets stemming from related studies.

used model is the multiple confirmatory factor analysis (MCFA) model proposed by Jöreskog (1971). The major limitation of Jöreskog (1971)'s model is that it assumes normality for the observed data. Although, recent review papers on the topic of applied statistics to medical researches call attention for the violation of the normality assumption in several situation commonly occurring in medical studies (MURPHY, 2004; GENSER *et al.*, 2007; WANG *et al.*, 2015).

FA models allowing for the relaxation of the normality assumption has appeared in the statistical literature since at least Browne and Shapiro (1987), with the author proposing the use of scale mixture of normal (SMN) distributions (ANDREWS and MALLOWS, 1974) for the common latent factors in the exploratory FA model. The SMN class of probability distributions includes, as special cases, the normal distribution and distributions with heavy tails as the t-Student and contaminated normal (WEST, 1987). Since it was proposed by Jöreskog (1971), the MCFA model have been extensively studied mainly in what concerns hypothesis testing of invariance using the likelihood ratio test (YUAN and BENTLER, 2004, 2006; YUAN and CHAN, 2016), but according to our literature review, estimation of MCFA models for non-normal responses appears only in the Bayesian statistical literature. Song and Lee (2001) proposed a Bayesian MCFA for handling mixed types of continuous and ordinal variables.

The importance of introducing MCFA models adequate for modeling non-normal response data relies in the increasing interest of medical researchers in testing new theories regarding gene functions in the human physiology and their interaction through co-expression for regulating biological pathways (CABRAL-MARQUES and RIEMEKASTEN, 2017). The SMN class of distributions proposed by Andrews and Mallows (1974) offers a theoretically sound framework for extending the MCFA model of Jöreskog (1971) to include latent factors with heavier tails than the normal distribution, hence allowing for more flexible data analysis.

In our research the main objective is to define and to estimate a new MCFA model integrating the SMN class of distributions in the probabilistic assumptions of the model. The usefulness of the new model shall be confirmed by means of Monte Carlo simulation studies and an application using real data stemming from researches in oncology. We propose and evaluate a sound hypothesis about a gene co-expression network regulating the pathology of pancreas cancer. Our hypothesis is based on a exploratory multiple group factor analysis and its validity is confirmed by comparison of the results with the specialized knowledge at disposal in the literature about the molecular biology of cancer (CASEY *et al.*, 2007; LI *et al.*, 2013; FANG *et al.*, 2014; GIALELI *et al.*, 2014; COX *et al.*, 2015; GASCARD and TLSTY, 2016; JIA *et al.*, 2016; HAMMER *et al.*, 2017).

1.3 Contributions

The main contributions of this research are the following:

- To extend the distributional assumptions of the MCFA model of Jöreskog (1971) by allowing the observed data to be distributed in the class of scale mixture of normal (SMN) distributions;
- To define identification conditions for the model's parameters;
- To develop an Expectation-Conditional-Maximization (ECM) algorithm for estimation of the model's parameters.

1.4 Preliminaries

In the following, we review the theory of factor analysis and present the SMN class of distributions.

1.4.1 Factor analysis

The factor analysis (FA) model presented by Jöreskog (1969) describes a p-dimensional random variable Y in terms of latent variables through the linear equation

$$Y = \mu + \Lambda Z + \varepsilon, \tag{1.1}$$

where μ, of order $p \times 1$, is an intercept, Λ is a $p \times k$ loading matrix, Z is a $k \times 1$ random vector of common latent factors that explain the shared variation of the p dimensions of Y and ε is a $p \times 1$ random vector of noise specific to each dimension of Y. In addition, suppose $(Z, \varepsilon)^\top$ follows a multivariate normal distribution given by

$$\begin{bmatrix} Z \\ \varepsilon \end{bmatrix} \sim \mathsf{N}_{p+k}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \zeta & 0 \\ 0 & \Psi \end{bmatrix} \right), \tag{1.2}$$

where ζ is a covariance matrix for the common latent factors and Ψ is diagonal variance matrix for the specific factors.

The FA model (1.1) with assumption (1.2) says that the variability common to all p dimensions of Y is completly determined by the common latent factors Z (RUBIN and THAYER, 1982). Also, its possible to demonstrate (ANDERSON and RUBIN, 1956) that the model induces the following non-linear latent structure on Σ, the covariance matrix of Y,

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\zeta}\boldsymbol{\Lambda}^{\top} + \boldsymbol{\Psi}, \tag{1.3}$$

which is known in the literature as the system of normal equations for the FA model (REILLY, 1995).

The FA model can be interpreted in two forms (BEKKER *et al.*, 1994, pp. 75-76). The first form of interpretation leads to the Confirmatory Factor Analysis (CFA), where the linear equation (1.1) is seen as a model of causation. In CFA, the common factors Z are underling unobserved variables generating Y, although random noise can affect each dimension of Y. Specialized knowledge about the underling process generating Y can be introduced in the CFA model as constraints in its parameters, specifically, by pre-specifying values to some elements of Λ, ζ or Ψ (JÖRESKOG, 1969). The second form of interpretation of FA is called Exploratory Factor Analysis (EFA). In EFA, the main objective is to represent the observed variables in a space of smaller dimension. That is accomplished by representing the observed variables as linear combinations of latent factors. Although, in EFA parameter constraints are included only to identify the model (JÖRESKOG, 1967).

The CFA and EFA models share the same basic model structure defined in (1.1) and (1.2), with their main difference residing on the matrix ζ and model identification strategy. In CFA models, ζ is conceptualized in a way that allows any kind of constraint in its parameters, among fixed values and equality constraints. Although, in EFA the covariance matrix ζ is fixed and equal to the identity matrix of order k. Regarding model identification, the EFA model is just identified, *i.e.* the number of identifying restrictions in the model's parameters are not greater than the number necessary for guaranteeing model identification, while in CFA models the parameters are usually over-identified, *i.e.* there exist more than one normal equation in the system (1.3) contributing to a unique solution to some of the parameters, as shown by Bollen (1989, pp. 88-89).

An important extension of the CFA model allows for the simultaneous analysis of independent populations. This model was proposed by Jöreskog (1971) and is called Multiple Group Confirmatory Factor Analysis (MCFA). The MCFA modeling context arises in situations where a researcher wants to test if a specified hypothetical latent structure could accurately describe the common variability of a set of variables observed in $G \geq 1$ independent groups of individuals. In a typical application in the field of Psychometrics, the MCFA model would allow the researcher to verify if a set of hypothetical constructs of his interest could be studied in two or more independent groups of individuals using the same measurement instrument in all groups (JÖRESKOG, 1971; SÖRBOM, 1974).

The MCFA model proposed by Jöreskog (1971) corresponds to $G \geq 1$ simultaneous CFA models, each one defined as in (1.1) and (1.2), but with the additional feature that parameters could be shared among groups. That new feature amounts to introduce a dependence of the parameter matrices in each of the G CFA models on a general vector of parameters θ. Hence, for $g \in \{1, \ldots, G\}$, the MCFA model is mathematically expressed as

$$Y_g = \mu_g + \mathbf{\Lambda}_g(\theta) Z_g + \varepsilon_g \tag{1.4}$$

and

$$\begin{bmatrix} Z_g \\ \varepsilon_g \end{bmatrix} \sim \mathsf{N}_{p_g+k_g}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \zeta_g(\theta) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}_g(\theta) \end{bmatrix} \right), \tag{1.5}$$

where p_g and k_g correspond to the dimensions of the random vectors Y_p and Z_g, respectively.

If in (1.4) it is supposed that the intercept μ_g also depends on θ and the expected value of Z_g in (1.5) is non-zero, $g = 1, \ldots, G$, then the model of Jöreskog (1971) extends to a more general version of the MCFA model proposed by Sörbom (1974). The models of Jöreskog (1971) and Sörbom (1974) differ not only in the estimation of parameters, but also and more decisively in its identifiability conditions. The field of application of both models differ according to the degree of invariance the researcher is interested in investigate.

Meredith and Teresi (2006) reviewed the definition and assumptions of factorial invariance, highlighting the most common models setups of MCFA used for research.

The hierarchy of factorial invariance starts with *configural invariance*, where its assumed that $\mathbf{\Lambda}_g(\theta)$ in (1.4) have the same configuration of fixed and free parameters, as well as the same number of latent common factors, in all G groups. When $\mathbf{\Lambda}_g(\theta)$ is exactly the same in all G groups we face the kind of invariance called *pattern invariance*. Together, configural and pattern invariance are considered weak forms of invariance, since they do not guarantee direct comparison between observed variables across groups. With weak invariance, the most the researcher can assert about the underlying latent structure relating the G groups is that the observed variables are measuring the same set of constructs in all groups. The remaining two levels of invariance are called *strong invariance* and *strict invariance*, and are both related to invariance of $\mu_g(\theta)$ and common factor means. Hence, the MCFA model defined in (1.4) and (1.5) can only account for configural and pattern invariance, although invariance of the matrices $\mathbf{\Psi}_g$, $g = 1, \ldots, G$, and partial levels of invariance obtained by restricting only specific parameters to be equal across groups are also allowed in MCFA (MEREDITH and TERESI, 2006).

1.4.2 Scale mixture of normal distributions

The scale mixture of normal (SMN) distributions was proposed by Andrews and Mallows (1974) in a study discussing the necessary and sufficient conditions for the existence of a random variable X generated as the ratio Z/U, where Z has a standard normal distribution and U is independent of Z. Andrews and Mallows (1974) determined the density function of X and established ways for determining the distribution of U.

In the multivariate case, a random vector X belonging to the SMN class of distributions can still be characterized by its stochastic representation, analogously as proposed by Andrews and Mallows (1974). The definition below gives the desired statement.

Definition 1. *A p-dimentional random vector X_p with location parameter μ and scale matrix Σ is in the SMN class of distributions if there exists a positive uni-dimensional random variable U, such that the following stochastic representation is valid*

$$X_p = \mu + \frac{Z_p}{\sqrt{U}}, \tag{1.6}$$

where $Z_p \sim N_p(\mathbf{0}, \Sigma)$ is distributed independently of U.

Following Andrews and Mallows (1974), an immediate result of Definition 1 is the density function f$(\cdot)$ of X_p, which is given by

$$\mathrm{f}(x) = |\Sigma|^{-1/2} \int_0^\infty \left(\frac{u}{2\pi}\right)^{p/2} \exp\left[-\frac{u}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right] d\mathrm{H}(u|\boldsymbol{\nu}), \tag{1.7}$$

where H$(\cdot|\boldsymbol{\nu})$ is the distribution function of U parametrized by the vector $\boldsymbol{\nu}$. From now on, we shall denote a random vector with the stochastic representation (1.6) or, equivalently, with density function (1.7) by $X_p \sim \mathrm{SMN}_p(\mu, \Sigma, \mathrm{H}(\cdot|\nu))$.

The work of Andrews and Mallows (1974) extended the results of an early work published by Beale and Mallows (1959) on the properties of scale mixing of symmetric distribution. Beale and Mallows (1959) had already proved several conditions on the moment of mixing distributions that allowed the conclusion that probability distributions in the SMN class have higher kurtosis than the normal distribution, except of course for the normal distribution itself. This fact is pointed out by Kano (1994).

Properties of the SMN distributions can be obtained by noting its relation to the elliptical class of distributions. Fang and Zhang (1990) give a full discussion of elliptical distributions and proves the following property, which holds for the SMN distributions.

Property 1. *Let $X_p \sim SMN_p(\mu, \Sigma, H(\cdot|\nu))$. For any matrix $\boldsymbol{A}$ of order $d \times p$ and of full rank, and for any vector α of dimension $d \times 1$, there is a random variable $Y_d = \alpha + \boldsymbol{A}X_p$ such that $Y_d \sim SMN_d\left(\alpha + \boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\Sigma\boldsymbol{A}^\top, H(\cdot|\nu)\right)$.*

Proof. The proof is in Fang and Zhang (1990, p. 66). □

Special cases of distribtuions in the SMN class were already given by Andrews and Mallows (1974), as for example, the t-Student, logistic and Laplace distributions. West (1984) and West (1987) gives other examples of distribtuions in the SMN class, as for example, the contaminated normal and power exponential distribution.

Next, we shall characterize the four types of SMN distributions explored in our research:

- Normal distribution: the normal distribution is obtained when in the stochastic representation (1.6) the random variable U is degenerated in 1, such that $P(U = 1) = 1$;

- t-Student distribution: when $U \sim \text{Gamma}(\nu/2,\nu/2)$, then the random variable X with stochastic representation (1.6) follows a p-variate t-Student distribution with location parameter μ and scale matrix Σ, denoted as $\text{t}_p(\mu,\Sigma,\nu)$. The density function is given by

$$\text{f}(x|\nu) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})\nu^{p/2}\pi^{p/2}|\Sigma|^{1/2}}\left[1+\frac{1}{\nu}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right]^{-\frac{\nu+p}{2}} \tag{1.8}$$

- Contaminated normal distribution: Let γ be real number in the interval $(0,1)$. When U is a discrete random variable with distribution given by $\text{P}(U=\gamma)=\xi$ and $\text{P}(U=1)=1-\xi$, then the random variable X with stochastic representation (1.6) follows a p-variate contaminated normal distribution, denoted as $\text{CN}_p(\mu,\Sigma,\gamma,\xi)$ with density given by

$$\text{f}(x|\xi,\gamma) = \xi\phi_p(x|\mu,\gamma^{-1}\Sigma)+(1-\xi)\phi_p(x|\mu,\Sigma), \tag{1.9}$$

  where $\phi_p(\cdot|\mu,\Sigma)$ is the density of a p-variate normal random variable with location μ and variance Σ,

- Slash distribution: when $U \sim \text{Beta}(\nu,1)$ then the random variable X with stochastic representation (1.6) follows a p-variate slash distribution, denoted as $\text{SL}_p(\mu,\Sigma,\nu)$. The density function is given by

$$\text{f}(x|\nu) = \nu^{p/2}\int_0^1 \frac{u^{\nu-1}}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{u}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right]du. \tag{1.10}$$

2 Multiple group factor analysis with SMN distributions

2.1 Resumo da seção

Iniciamos nesta seção a teoria do modelo de análise fatorial confirmatório em múltiplos grupos supondo distribuição dos fatores latentes na classe SMN. Denotamos o modelo por MCFA-SMN. O modelo é definido e sua relação com outros modelos presentes na literatura é estabelicida, destacando-se casos particulares. A função de verossimilhança é apresentada, justificando a necessidade de um algoritmo de estimação alternativo, o algoritmo ECM. A identificabilidade do modelo é tratada de forma geral, com a apresentação de condiçoes necessárias e/ou suficientes para a identificação dos parâmetros. A subseção onde se trata da estimação dos pararâmetros do modelo apresenta um algoritmo ECM que cumpre a propriedade *space filling* de Meng and Rubin (1993), garantindo as propriedades de convergência do algoritmo. Métodos para estimar o desvio padrão das estimativas de máxima verossimilhança são apresentados. Estes métodos dispensam o cáculos de derivadas de segunda ordem.

2.2 Model definition

Suppose a CFA model holds in each of G distinct groups or populations. The individual CFA models shall be called sub-models. Additionally, suppose there exist the prior knowledge that sub-models could share parameters with each other in a well defined way. Let θ be a generic vector comprised by the parameters existent in all sub-models, except possibly by parameter intercepts. Mathematically, the MCFA-SMN model is specified in terms of the random vector Y_{ig} of order $p_g \times 1$ related to latent factors through the equation

$$Y_{ig} - \mu_g = \mathbf{\Lambda}_g(\theta) Z_{ig} + \varepsilon_{ig},\ g = 1, \ldots, G, \tag{2.1}$$

where $i = 1, \ldots, n_g$ is a subject index, μ_g is a $p_g \times 1$ intercept specific for the g-th group, $\mathbf{\Lambda}_g(\theta)$ is a $p_g \times k_g$ matrix of loading coefficients dependent on the vector of parameters θ, Z_{ig} is a $k_g \times 1$ random vector of common latent factors and ε_{ig} is a $p_g \times 1$ random vector of specific latent factors (also called errors).

In this case, we consider that vectors of common and specific latent factors are jointly distributed as

$$\begin{bmatrix} Z_{ig} \\ \varepsilon_{ig} \end{bmatrix} \sim \mathsf{SMN}_{p_g+k_g}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \zeta_g(\theta) & 0 \\ 0 & \Psi_g(\theta) \end{bmatrix}, \nu \right), \; g = 1, \ldots, G, \tag{2.2}$$

where $\zeta_g(\theta)$ is the covariance matrix of Z_{ig} and $\Psi_g(\theta)$ is the diagonal variance matrix for ε_{ig}, both dependent on the vector of parameters θ. The parameter vector ν indexes the common distribution of the mixing variables $U_{ig} \sim \mathsf{H}(\cdot|\nu)$ that defines the SMN distribution in (2.2), $g = 1, \ldots, G$. We assume the true value of ν is known and equal between the G groups.

Without loss of generality, for $g \in \{1, \ldots, G\}$, suppose Y_{ig} is corrected by its mean, μ_g. Hence, Equations (2.1) and (2.2) assert the random vector Y_{ig} is an affine combination of common, Z_{ig}, and specific, ε_{ig}, latent factors following a SMN distribution. Hence, by an application of Property 1, the MCFA-SMN model could be directly specified as

$$Y_{ig} \sim \mathsf{SMN}_{p_g}(0, \Sigma_g(\theta), \mathsf{H}(\cdot|\nu)), \; g = 1, \ldots, G, \tag{2.3}$$

where the scale matrix $\Sigma_g(\theta)$ has the latent structure

$$\Sigma_g(\theta) = \Lambda_g(\theta)\zeta_g(\theta)\Lambda_g^\top(\theta) + \Psi_g(\theta). \tag{2.4}$$

Based on (2.3) and in the density function (1.7), for a given random sample of size $n = \sum_{g=1}^{G} n_g$ taken from the G groups, $y = (y_{11}^\top, \ldots, y_{n_1 1}^\top, \ldots, y_{1G}^\top, \ldots, y_{n_G G}^\top)^\top$, the log-likelihood has the following form

$$\ell(\theta) = -\frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{n_g} \log|\Sigma_g(\theta)| + \sum_{g=1}^{G}\sum_{i=1}^{n_g} \log \int_0^\infty \mathsf{f}(y_{ig}|\theta) d\mathsf{H}(u|\nu), \tag{2.5}$$

where

$$\mathsf{f}(y_{ig}|\theta) = \left(\frac{u}{2\pi}\right)^{p_g/2} \exp\left(-\frac{u}{2} y_{ig}^\top \Sigma_g^{-1}(\theta) y_{ig}\right). \tag{2.6}$$

The estimation of θ can be achieved by direct maximization of (2.5) using quasi-Newton methods. For example, under the supposition of normality of latent factors Equation (2.5) reduces to the log-likelihood of Jöreskog (1971), whose developed a modified Fletcher-Powell algorithm for maximum likelihood of θ.

Further specification of the invariance pattern of θ can lead to parsimonious model structures, including all models of interest present in the literature. We shall demonstrate through three examples how a set of restrictions placed on θ transforms the MCFA-SMN models into some important MCFA models. For now, suppose the observable variables Y_{ig} follow a normal distribution, $g = 1, \ldots, G$, and denote the resultant model as MCFA-N.

Example 1. *If $\theta = (\theta_1, \ldots, \theta_g)$ and the dependence of the model matrices on θ is expressed as $\mathbf{\Lambda}_g(\theta) = \mathbf{\Lambda}_g(\theta_g)$, $\boldsymbol{\zeta}_g(\theta) = \boldsymbol{\zeta}_g(\theta_g)$ and $\mathbf{\Psi}_g(\theta) = \mathbf{\Psi}_g(\theta_g)$, for $g = 1, \ldots, G$, then the MCFA-N model corresponds to a separate CFA model for G populations. If $G = 1$, the CFA model of Jöreskog (1969) is recovered.*

Example 2. *If in Example 1, $p_g = p$ and $k_g = k$, for all $g = 1, \ldots, G$, and the dependence of the loading matrices on θ induces the partition $\mathbf{\Lambda}_g(\theta) = \begin{bmatrix} \mathbf{\Lambda}^{(1)} & \mathbf{\Lambda}_g^{(2)} \end{bmatrix}$, for $g = 1, \ldots, G$, where $\mathbf{\Lambda}^{(1)}$ is shared among all G groups and $\mathbf{\Lambda}_g^{(2)}$ is specific to the g-th group, then the MCFA-N model corresponds to a confirmatory version of the model proposed by De Vito (2016). Additionally, if the covariance matrix of common latent factors does not depend on θ and $\boldsymbol{\zeta}_g(\theta) = I_k$, for all $g = 1, \ldots, G$, then the same model as proposed by De Vito (2016) is recovered.*

Example 3. *If in Example 2, $\mathbf{\Lambda}_g^{(2)} = \mathbf{\Lambda}^{(2)}$ or, equivalently, if $\mathbf{\Lambda}_g(\theta) = \mathbf{\Lambda}$, for all $g = 1, \ldots, G$, then the resultant model is a MCFA with pattern invariance. That kind of model was studied by (JÖRESKOG, 1971), and is recovered by the MCFA-N model, as it should be.*

In assuming a SMN distribution for the vector of latent factors the MCFA-SMN model generalizes the model of Jöreskog (1971), which restricts the vector of latent factors to have a multivariate normal distribution. The model of De Vito (2016) is a particular case of Jöreskog (1971), consequently being also generalized by the MCFA-SMN model. Although still under the assumption of multivariate normality for the latent factors, the model of Sörbom (1974) is not a particular case of the MCFA-SMN model, since the Sörbom (1974)'s model allows the intercepts μ_g to be equal among the G groups and also allows the latent common factors to have non-zero means.

From now on, we will omit the dependence of model matrices on the vector of parameters θ in situations where it will not cause any confusion. Hence, for any $g \in$

$\{1,\ldots,G\}$, we shall denote $\boldsymbol{\Lambda}_g = \boldsymbol{\Lambda}_g(\theta)$, $\boldsymbol{\zeta}_g = \boldsymbol{\zeta}_g(\theta)$ and $\boldsymbol{\Psi}_g = \boldsymbol{\Psi}_g(\theta)$, unless it is stated differently.

2.3 Identifiability

Most factor analysis models share two sources of indeterminacy in its parameters, which became known in the literature as the uniqueness and identification problems (BOLLEN and JÖRESKOG, 1985). The uniqueness problem stems from the invariance of factor loading matrices under post-multiplication by an unrestricted non-singular matrix, although in the EFA model that matrix is necessarily orthogonal (ANDERSON and RUBIN, 1956). In the context of the CFA model, that sort of invariance is related to rotation, reflexion and sign changes in the common factors' covariance matrix and, as well as in EFA, it is a source of non-identifiability of parameters. In the following, we adapted the definition of the uniqueness problem given by Bollen and Jöreskog (1985) to the context of MCFA-SMN models.

Definition 2. *In the MCFA-SMN model, the uniqueness of $\boldsymbol{\Lambda}_g$ and $\boldsymbol{\zeta}_g$ are established if for every non-singular matrix $\boldsymbol{T}_g$ the transformations $\boldsymbol{\Lambda}_g^* = \boldsymbol{\Lambda}_g \boldsymbol{T}_g^{-1}$ and $\boldsymbol{\zeta}_g^* = \boldsymbol{T}_g \boldsymbol{\zeta}_g \boldsymbol{T}_g^\top$ do not change any of the constrained or unconstrained entries of $\boldsymbol{\Lambda}_g$ or $\boldsymbol{\zeta}_g$, $g = 1,\ldots,G$.*

For the MCFA-SMN model, the uniqueness problem posits the existence of observationally equivalent parameter vectors $\theta_1 = (\boldsymbol{\Lambda}_g, \boldsymbol{\zeta}_g)_{g=1}^G$ and $\theta_2 = (\boldsymbol{\Lambda}_g^*, \boldsymbol{\zeta}_g^*)_{g=1}^G$, *i.e.* $\theta_1 \neq \theta_2$ implying $\ell(\theta_1) = \ell(\theta_2)$, where $\ell(\cdot)$ is the model's log-likelihood function for a given sample of observations. Hence the uniqueness problem imposes a search for $\boldsymbol{\Lambda}_g$ and $\boldsymbol{\zeta}_g$ such that $\boldsymbol{\Lambda}_g^* = \boldsymbol{\Lambda}_g \boldsymbol{T}_g^{-1}$ and $\boldsymbol{\zeta}_g^* = \boldsymbol{T}_g \boldsymbol{\zeta}_g \boldsymbol{T}_g^\top$ imply $\boldsymbol{T}_g = \boldsymbol{I}$, $g = 1,\ldots,G$.

We shall broadly state the identification problem of statistical models in general and subsequently present it in terms of the MCFA-SMN model through a definition. In a general modeling framework, the local identification of a parameter vector $\theta_1 \in \mathscr{H}$ is attained when a neighborhood of θ_1 has no other vector θ_2 that is observationally equivalent to θ_1, unless $\theta_1 = \theta_2$ (BEKKER *et al.*, 1994, pp. 17-18). If that neighborhood of θ_1 coincides with $\mathscr{H}$, then θ_1 is said to be globally identified in $\mathscr{H}$ (BEKKER *et al.*, 1994, pp. 19). Partial identification of θ is attained when some, but not all, of its entries are identified (BEKKER *et al.*, 1994, pp. 17). To define the identification problem for MCFA-SMN models we adapted a statement of Reilly (1995) associating the identification of parameters in CFA models to its normal equations.

Definition 3. *Let $\theta \in \mathscr{H} \subset \mathbb{R}^d$ be the vector having as its elements all distinct parameters that composes the set of matrices $\{\boldsymbol{\Lambda}_g, \boldsymbol{\zeta}_g, \boldsymbol{\Psi}_g\}_{g=1}^{G}$ in the MCFA-SMN model. The vector of parameters θ is identified in $\mathscr{H}$ if it is uniquely determined by $\boldsymbol{\Sigma}_{(g)}$, $g = 1, \ldots, G$, through the system of normal equations (2.4).*

According to Definition 3, identification of parameters in the MCFA-SMN model depends at most on second-order moments conditions. Indeed, that is true for such models arising from Equations (2.1) and (2.2) when the parameter vector ν indexing the distribution of the mixing variable is known. Bentler (1983) states that the fourth-order moments of elliptical distributions are exclusively expressed in terms of second-order moments and a kurtosis parameter κ, that in the case of SMN distributions is in turn a function of ν. Hence, in the MCFA-SMN model if ν is free for estimation its identification may depend on additional conditions aside from that ones presented in Definition 3. Although not mentioned in the Definition 3, the parameters μ_g, $g = 1, \ldots, G$, in the MCFA-SMN model are clearly identified from the first-order moments. Nonetheless, if μ_g, as a function of θ, could share some or all of its entries between a subset $\mathscr{B} \subset \{1, \ldots, G\}$ of groups, that would be necessary extra conditions based on first-order moments for the global identification of θ (SÖRBOM, 1974).

Bollen and Jöreskog (1985) give an example of CFA model with parameters specified in such a way that uniqueness of the factor loading and latent factors' covariance matrices does not lead to identification of the whole model's parameters, showing empirically that the uniqueness and identification problems are not equivalents. In that same context, Peeters (2012) emphasizes that identification of the factor loading matrix may depend on the identification of specific errors' variance matrix. In the sequel we shall give rules for the solution of identification problems in the MCFA-SMN model, although we leave the uniqueness problem aside.

The following theorem formalizes a simple rationale behind parameter identification in multiple group factor analysis models in general and that permeates most of its practical applications (JÖRESKOG, 1971; SÖRBOM, 1974; SONG and LEE, 2001; DE VITO, 2016). We refer to the practice of asserting the identification of θ when each of the vectors $\theta_g = (\boldsymbol{\Lambda}_g(\theta), \boldsymbol{\zeta}_g(\theta), \boldsymbol{\Psi}_g(\theta))$ are simultaneously identified, $g = 1, \ldots, G$.

Theorem 1. *In the MCFA-SMN model, let $\theta_g = (\boldsymbol{\Lambda}_g(\theta), \boldsymbol{\zeta}_g(\theta), \boldsymbol{\Psi}_g(\theta))$ and $\theta = (\theta_1, \ldots, \theta_G)$. If $\theta_g \in \mathscr{H}$ is identified for all $g \in \{1, \ldots, G\}$, then θ is identified in $\mathscr{H}$.*

Proof. Using the definition of parameter identification given by Bekker, Merckens, and Wansbeek (1994, pp. 19), identification of θ is achieved if and only if in a neighborhood of θ each of its parts θ_g, $g = 1, \ldots, G$, is identified. Hence, considering a neighborhood of $\theta \in \mathscr{H}$ where it is locally identified, in this neighborhood all the components θ_g, $g = 1, \ldots, G$, are also locally identified. Conversely, if there exist a neighborhood where $\theta_g \in \mathscr{H}$ is locally identified, for all $g = 1, \ldots, G$, then θ is identified in this neighborhood. For attaining global identification of θ in $\mathscr{H}$, the neighborhood defined in the previous statements must coincide with $\mathscr{H}$. $\square$

Theorem 1 gives sufficient conditions for the identification of θ in MCFA-SMN models with the global or local status of its identification depending on the degree of identification of its components θ_g, $g = 1, \ldots, G$. Since the theorem uses the identification of θ_g, $g = 1, \ldots, G$, as a way to identify θ, it reduces the identification problem in MCFA-SMN models to the problem of identifying parameters in G independent CFA models. Hence, the contribution of Theorem 1 is making available for the identification of MCFA-SMN models all the well known rules of identification for a single factor analysis model at disposal in the literature (ANDERSON and RUBIN, 1956; BOLLEN, 1989; REILLY, 1995; REILLY and O'BRIEN, 1996; GEWEKE and ZHOU, 1996; BEKKER and TEN BERGE, 1997; BAI and LI, 2012; PEETERS, 2012, among others).

An important example of identification constraint appears in applications of exploratory factor analysis models. To define the desired set of constraints, we follow Bai and Li (2012) and define recursively the $G \geq 1$ loading matrices entering in the MCFA-SMN model. The loading matrix $\boldsymbol{\Lambda}_g$ will be such that its first column has only non-zero loadings, while in the second column it has the loading $\lambda_{1,2}^{(g)} = 0$, in the third column $\lambda_{1,3}^{(g)} = \lambda_{2,3}^{(g)} = 0$, and so on until the k_g-th column where $\lambda_{1,k_g}^{(g)} = \cdots = \lambda_{k_g-1,k_g}^{(g)} = 0$, for $g = 1, \ldots, G$. That guarantees a $\boldsymbol{\Lambda}_g$ partitioned as an lower triangular matrix of dimensions $k_g \times k_g$ and an unrestricted matrix of dimension $(p - k_g) \times k_g$. Anderson and Rubin (1956) call this type of constraint the *triangular matrix of zeros*.

A practical example of an application of this kind of constraint shall be explored in Section 4, where the reader can find an explicit presentation of loading matrices following a *triangular matrix of zeros* form. The next proposition shall give the sufficient conditions for parameter identification in MCFA-SMN models defined with loading matrices following a *triangular matrix of zeros* form.

Proposition 1. *Define a MCFA-SMN model where the $G \geq 1$ loading matrices entering in the model are all in a triangular matrix of zeros form, as defined in Anderson and Rubin (1956), and the covariance matrices of common factors are identity matrices of appropriate order. Then if the loading parameters $\lambda_{j,j}^{(g)}$, $j = 1,\ldots,k_g$, and $g = 1,\ldots,G$, are identified, then all the parameters in the MFCA-SMN model are also identified.*

Proof. From the normal equations (2.4), it can be seen that the covariance matrix $\boldsymbol{\Sigma}_g = \left(\sigma_{i,j}^{(g)}\right)$ has its typical element of form

$$\sigma_{i,j}^{(g)} = \sum_{x=1}^{k_g}\sum_{y=1}^{k_g} \lambda_{i,x}^{(g)}\lambda_{j,y}^{(g)}\varsigma_{x,y}^{(g)} + \psi_{i,j}^{(g)}, \tag{2.7}$$

where $\lambda_{i,j}^{(g)}$, $\varsigma_{i,j}^{(g)}$ and $\psi_{i,j}^{(g)}$ are the elements in the i-th row and j-th column of the matrices $\boldsymbol{\Lambda}_g$, $\boldsymbol{\zeta}_g$ and $\boldsymbol{\Psi}_g$, respectively.

Constraining the MCFA model to have loading matrices following a *triangular matrix of zeros* form and common factors with variances equal to unite and covariances equal to zero, the implied normal equations have the following general form

$$\sigma_{i,j}^{(g)} = \sum_{k=1}^{i} \lambda_{i,k}^{(g)}\lambda_{j,k}^{(g)} + \psi_{i,j}^{(g)},\ i = 1,\ldots,p_g \text{ and } j = 1,\ldots,p_g. \tag{2.8}$$

Assume $\lambda_{j,j}^{(g)}$, $j = 1,\ldots,k_g$, and $g = 1,\ldots,G$, are identified. With this restriction the identification of $\psi_{j,j}^{(g)}$, $j = 1,\ldots,p_g$, is immediate. Now, the solution to (2.8) in terms of the remaining parameters is recursive. Consider g fixed. Since for $i \neq j$ $\psi_{i,j}^{(g)}$ equals to zero, we conclude that $\lambda_{j,1}^{(g)}$ is identified, for all $j = 2,\ldots,p_g$. Now the general rule is that identification of any $\lambda_{i,j}^{(g)}$, with $i \neq j$ and both different from unite, follows from the simultaneous identification of $\lambda_{i-1,j}^{(g)}$ and $\lambda_{i,j-1}^{(g)}$. Hence, identification of $\lambda_{j,1}^{(g)}$ together with the identifiability assumption of $\lambda_{j,j}^{(g)}$, for all $j = 1,\ldots,p_g$, imply identification of the parameters in the matrix $\boldsymbol{\Lambda}_g$. That way, for a fixed g, $\boldsymbol{\Lambda}_g$ and $\boldsymbol{\Psi}_g$ have its parameters identified. To extend identification to the entire MCFA-SMN model we use the results of Theorem 1, since according to this theorem the MCFA-SMN model will be fully identified if each of its G parts is simultaneously identified. $\square$

To understand the Theorem 1 we could begin with the basic modeling framework of a separate CFA model for G groups, as defined in Example 1 of Section 2.2. In that case, the parameter vectors $\theta_g = (\mathbf{\Lambda}_g, \zeta_g, \mathbf{\Psi}_g)$, $g = 1, \ldots, G$, partitions θ and the conditions stated in Theorem 1 become not only sufficient but also necessary for parameter identification. Certainly, any other possible set up for a multiple group CFA model must derive from the separate CFA model for G groups through the specification of relationships between parameters. According to Reilly (1995), equality constraint should not hinder the identification of θ, since it can only reduce the number of available solutions of the original unconstrained model.

Parameters are said over-identified when there exist more then one normal equation in the system (2.4) establishing the identification of the parameter. Although, in well specified models all the normal equations should determine a unique solution involving the parameter (BOLLEN, 1989, pp. 89-90). It is also important to notice that, differently from equality between parameters, constraints on the form of fixed values for a set of parameters could turn an identified model into an unidentified model (REILLY, 1995). The next example illustrates this last fact and shows the conditions of Theorem 1 are only sufficient for parameter identification of MCFA-SMN models.

Example 4. *Consider the following set up of parameter matrices for a MCFA-SMN model for $G = 2$ groups:*

$$\mathbf{\Lambda}_g(\theta) = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ 0 & \lambda_3 \\ 0 & \lambda_4 \end{bmatrix}, \; \mathbf{\Psi}_g(\theta) = \begin{bmatrix} \psi_1^{(g)} & 0 & 0 & 0 \\ 0 & \psi_2^{(g)} & 0 & 0 \\ 0 & 0 & \psi_3^{(g)} & 0 \\ 0 & 0 & 0 & \psi_4^{(g)} \end{bmatrix}, \; \zeta_g(\theta) = \begin{bmatrix} 1 & \varsigma_{1,2}^{(g)} \\ \varsigma_{1,2}^{(g)} & 1 \end{bmatrix}, \; g = 1{,}2.$$

The parameter vectors defined in Theorem 1 are $\theta_g = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \psi_1^{(g)}, \psi_2^{(g)}, \psi_3^{(g)}, \psi_4^{(g)}, \varsigma_{1,2}^{(g)})$, $g = 1,2$ and $\theta = (\theta_1, \theta_2)$. Reilly (1995) showed that for any sub-model θ_g is identified if and only if $\varsigma_{1,2}^{(g)} \neq 0$, $g = 1,2$. If this condition is satisfied, then, by Theorem 1, we conclude θ is identified. In another situation, suppose $\varsigma_{1,2}^{(2)} = 0$. Now θ_2 does not fulfill the identification condition of Reilly (1995), hence Theorem 1 can not be used to establish the identification of θ. Yet, since θ_1 remains identified and $\zeta_2(\theta)$ is fixed, all distinct parameters that comprise θ are identified.

A solution for the system of equations defined by the normal equations presented in (2.7) depends on constrained parameters. As presented by Anderson and Rubin

(1956), another useful set of constraints known as *simple structure* establishes that each observed variable loads in exactly one latent variable . Hence, the *simple structure* imposes a loading matrix with each row being composed of only zeros except in one of its entries. Example of models with parameters following a *simple structure* are any of the two parts $(g = 1, 2)$ of the MCFA-SMN model presented in Example 4. That same example allows to conclude that, at least for fixed g, *simple structure* does not guarantees parameter identification.

In a MCFA-SMN model, $\theta = (\theta_1, \ldots, \theta_G)$ is under a *simple structure* if all of its G components simultaneously follow a *simple structure*. In this case, Equation (2.7) simplifies to

$$\sigma_{i,j}^{(g)} = \lambda_{i,x^*}^{(g)} \lambda_{j,y^*}^{(g)} \varsigma_{x^*,y^*}^{(g)} + \psi_{i,j}^{(g)}, \; g = 1, \ldots, G, \tag{2.9}$$

with x^* and y^* specifying the columns of the $\mathbf{\Lambda}_g$ matrix where the i-th and j-th rows, respectively, have a non-zero loading. Also, x^* and y^* are dependent on the group index g. Reilly (1995) devised a necessary and sufficient rule, based on inspection of equation (2.9), for identification of parameters in models for a single group $(G = 1)$ under *simple structure*. We shall present a theorem extending Reilly (1995)'s identification rule to models with $G \geq 1$.

The following theorem generalizes to the context of MCFA-SMN models the Proposition 1 of Reilly (1995). The Proposition 1 of Reilly (1995) appears as a particular case when $G = 1$ in the MCFA-SMN model. The proof given by Reilly (1995) remain valid under minor modifications. Hence the proposed theorem should be viewed as a new perspective of Reilly (1995)'s result. Moreover, the conditions appearing in the theorem guarantees global identification of θ (REILLY, 1995). Hence, the following theorem should represent an important advance in the study of identifiability in multiple group factor analysis models.

Theorem 2. *Consider a MCFA-SMN model under simple structure and indexed by the parameter vector θ. Define $\mathscr{P} = \left\{ \sigma_{i,j}^{(g)} = \lambda_{i,x^*}^{(g)} \lambda_{j,y^*}^{(g)} \varsigma_{x^*,y^*}^{(g)} + \psi_{i,j}^{(g)} \,\middle|\, \psi_{i,j}^{(g)} = 0, \; g = 1, \ldots, G \right\}$, the set of normal equations corresponding to a non-diagonal element of $\Sigma_g(\theta)$, $g = 1, \ldots, G$. Enumerate the $N = \sum_{g=1}^{G} p_g(p_g - 1)/2$ distinct elements of $\mathscr{P}$ and denote them as $\sigma_1, \ldots, \sigma_N$. Let $\sigma = (\sigma_1, \ldots, \sigma_N)$. The parameter vector θ is identified if and only if the*

Jacobian matrix

$$R_G = \frac{\partial \log|\sigma|}{\partial \log|\theta^\top|} \tag{2.10}$$

is of full column rank.

Proof. According to Proposition 1 of Reilly (1995), $\mathscr{P}$ is identified with at least one solution if and only if $|\mathscr{P}| = \left\{|\sigma_{i,j}^{(g)}| = |\lambda_{i,x^*}^{(g)}||\lambda_{j,y^*}^{(g)}||\varsigma_{x^*,y^*}^{(g)}| + \psi_{i,j}^{(g)} \,\middle|\, \psi_{i,j}^{(g)} = 0,\ g = 1,\ldots,G\right\}$ is also identified. Applying this result, the identification of $\mathscr{P}$ corresponds to a solution of the system of equations

$$\log\left|\sigma_{i,j}^{(g)}\right| = \log\left|\lambda_{i,x^*}^{(g)}\right| + \log\left|\lambda_{j,y^*}^{(g)}\right| + \log\left|\varsigma_{x^*,y^*}^{(g)}\right|, \tag{2.11}$$

where $g = 1,\ldots,G$ and $i,j = 1,\ldots,p_g$.

The system (2.11) can be written in matrix form as

$$\log|\sigma| = R_G \log|\theta|, \tag{2.12}$$

where R_G is a binary matrix that corresponds to the Jacobian (2.10).

Applying the chain rule,

$$R_G = \frac{\partial \log|\sigma|}{\partial \log|\theta^\top|} = \frac{\partial \log|\sigma|}{\partial \sigma}\frac{\partial \sigma}{\partial \theta^\top}\frac{\partial \theta}{\partial \log|\theta^\top|}.$$

The matrices $\frac{\partial \log|\sigma|}{\partial \sigma}$ and $\frac{\partial \theta}{\partial \log|\theta^\top|}$ are diagonal and of full column rank, hence rank$(R_G) =$ rank$\left(\frac{\partial \sigma}{\partial \theta^\top}\right)$. Following Reilly (1995), the Implicit Function Theorem guarantees a unique solution to the system (2.12) if and only if R_G is of full column rank.

Assuming R_G of full column rank, the set of parameters $\{\mathbf{\Lambda}_g(\theta), \zeta_g(\theta)\}_{g=1}^G$ is identified. Using similar arguments to Reilly (1995), the identification of θ is accomplished by noting that

$$\mathbf{\Psi}_g(\theta) = \mathbf{\Sigma}(\theta)_g - \mathbf{\Lambda}_g(\theta)\boldsymbol{\zeta}_g(\theta)\mathbf{\Lambda}_g^\top(\theta),\ g = 1,\ldots,G,$$

which shows that $\{\mathbf{\Psi}_g(\theta)\}_{g=1}^G$ is identified. $\square$

The matrix R_G defined in (2.10) is easy to be constructed and its construction follows the same description as given by Reilly (1995), but based in the new set $\mathscr{P}$ defined in Theorem 2. R_G has at most $N = \sum_{g=1}^G p_g(p_g-1)/2$ rows, each one corresponding to a normal equation selected from $\mathscr{P}$, and number of columns equals to the total number of unknown parameters involved in all selected equations. Consider the enumeration of the elements of $\mathscr{P}$ constructed in Theorem 2. The n-th row of R_G has either 0 or 1 entries, with a 1 placed in the columns corresponding to unknown parameters of the n-th normal equation enumerated from $\mathscr{P}$ and 0 in the remaining columns. Observe that $\mathscr{P}$ could have elements corresponding to redundant equations of form $\sigma_{i,j}^{(g)} = 0$. Those redundant equations result in null rows in the R_G matrix, so they could be omitted without changing the column rank of R_G.

We shall illustrate the usefulness of Theorem 2 by means of two examples. The first example recapitulates Example 4 and the second is an example elaborated to show how global identification of parameters in a MCFA-SMN model can be achieved starting only from underidentified sub-models entering in its composition.

Example 5. *Recapitulate the second situation of Example 4, where* $\varsigma_{1,2}^{(1)} \neq 0$ *and* $\varsigma_{1,2}^{(2)} = 0$*. The non-redundant normal equations in* $\mathscr{P}$ *and its associated matrix* R_2 *are*

$$\begin{aligned}
\sigma_{1,2}^{(1)} &= \lambda_1\lambda_2 & \sigma_{1,2}^{(2)} &= \lambda_1\lambda_2\\
\sigma_{1,3}^{(1)} &= \lambda_1\lambda_3\varsigma_{1,2}^{(1)} & \sigma_{3,4}^{(2)} &= \lambda_3\lambda_4\\
\sigma_{1,4}^{(1)} &= \lambda_1\lambda_4\varsigma_{1,2}^{(1)} & &\\
\sigma_{2,3}^{(1)} &= \lambda_2\lambda_3\varsigma_{1,2}^{(1)} & &\\
\sigma_{2,4}^{(1)} &= \lambda_2\lambda_4\varsigma_{1,2}^{(1)} & &\\
\sigma_{3,4}^{(1)} &= \lambda_3\lambda_4 & &
\end{aligned}
\qquad
R_2 = \begin{array}{c} \\ \sigma_{1,2}^{(1)} \\ \sigma_{1,3}^{(1)} \\ \sigma_{1,4}^{(1)} \\ \sigma_{2,3}^{(1)} \\ \sigma_{2,4}^{(1)} \\ \sigma_{3,4}^{(1)} \\ \sigma_{1,2}^{(2)} \\ \sigma_{3,4}^{(2)} \end{array}
\begin{array}{c}
\begin{array}{ccccc} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \varsigma_{1,2}^{(1)} \end{array}\\
\begin{bmatrix}
1 & 1 & 0 & 0 & 0\\
1 & 0 & 1 & 0 & 1\\
1 & 0 & 0 & 1 & 1\\
0 & 1 & 1 & 0 & 1\\
0 & 1 & 0 & 1 & 1\\
0 & 0 & 1 & 1 & 0\\
1 & 1 & 0 & 0 & 0\\
0 & 0 & 1 & 1 & 0
\end{bmatrix}
\end{array}$$

The 7*-th and* 8*-th rows of* R_2 *could be omitted since they are equal to the* 1*-th and* 2*-th rows, respectively. It implies that the entire identification of the model can reside only on the normal equations of* $\mathscr{P}$ *associated to* Σ_1*, and could be confirmed by the rank rule of Reilly (1995) applied over the first group* $g = 1$*. Of course, the resultant rank should be the same as rank*$(R_2) = 5$*. Since* R_2 *has full column rank,* $\theta = \left(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \varsigma_{1,2}^{(1)}\right)$ *is globally identified.*

Example 6. *Consider the following MCFA-SMN model under simple structure for* $G = 2$

groups:

$$g=1:\ \mathbf{\Lambda}_1(\theta)=\begin{bmatrix}\lambda_1 & 0\\ \lambda_2 & 0\\ 0 & \lambda_3\\ 0 & \lambda_4\end{bmatrix},\ \mathbf{\Psi}_1(\theta)=\begin{bmatrix}\psi_1^{(1)} & 0 & 0 & 0\\ 0 & \psi_2^{(1)} & 0 & 0\\ 0 & 0 & \psi_3^{(1)} & 0\\ 0 & 0 & 0 & \psi_4^{(1)}\end{bmatrix},\ \zeta_1(\theta)=\begin{bmatrix}\varsigma_{1,1} & \varsigma_{1,2}\\ \varsigma_{1,2} & \varsigma_{2,2}\end{bmatrix}$$

$$g=2:\ \mathbf{\Lambda}_2(\theta)=\begin{bmatrix}\lambda_1 & 0\\ \lambda_2 & 0\\ 0 & \lambda_3\\ 0 & \lambda_4\end{bmatrix},\ \mathbf{\Psi}_2(\theta)=\begin{bmatrix}\psi_1^{(2)} & 0 & 0 & 0\\ 0 & \psi_2^{(2)} & 0 & 0\\ 0 & 0 & \psi_3^{(2)} & 0\\ 0 & 0 & 0 & \psi_4^{(2)}\end{bmatrix},\ \zeta_2(\theta)=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

The non-redundant normal equations in $\mathscr{P}$ and its associated matrix R_2 are

$$\begin{aligned}
\sigma_{1,2}^{(1)} &= \lambda_1\lambda_2\varsigma_{1,1} & \sigma_{1,2}^{(2)} &= \lambda_1\lambda_2\\
\sigma_{1,3}^{(1)} &= \lambda_1\lambda_3\varsigma_{1,2} & \sigma_{3,4}^{(2)} &= \lambda_3\lambda_4\\
\sigma_{1,4}^{(1)} &= \lambda_1\lambda_4\varsigma_{1,2} & &\\
\sigma_{2,3}^{(1)} &= \lambda_2\lambda_3\varsigma_{1,2} & &\\
\sigma_{2,4}^{(1)} &= \lambda_2\lambda_4\varsigma_{1,2} & &\\
\sigma_{3,4}^{(1)} &= \lambda_3\lambda_4\varsigma_{2,2} & &
\end{aligned}$$

$$R_2=\begin{array}{c}
\\ \sigma_{1,2}^{(1)}\\ \sigma_{1,3}^{(1)}\\ \sigma_{1,4}^{(1)}\\ \sigma_{2,3}^{(1)}\\ \sigma_{2,4}^{(1)}\\ \sigma_{3,4}^{(1)}\\ \sigma_{1,2}^{(2)}\\ \sigma_{3,4}^{(2)}
\end{array}
\begin{array}{c}
\begin{array}{ccccccc}\lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \varsigma_{11} & \varsigma_{12} & \varsigma_{22}\end{array}\\
\begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 0\\
1 & 0 & 1 & 0 & 0 & 1 & 0\\
1 & 0 & 0 & 1 & 0 & 1 & 0\\
0 & 1 & 1 & 0 & 0 & 1 & 0\\
0 & 1 & 0 & 1 & 0 & 1 & 0\\
0 & 0 & 1 & 1 & 0 & 0 & 1\\
1 & 1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 1 & 0 & 0 & 0
\end{bmatrix}
\end{array}$$

Using the information at disposal in Example 4, we conclude the parameters for the separate model, i.e. holding g fixed, are underidentified. Reilly (1995)'s rank rule could be used to obtain the same conclusion for the separate model. Otherwise, considering the normal equations in $\mathscr{P}$ altogether leads to an R_2 matrix of rank 7. Hence, R_2 is of full rank and an application of Theorem 2 shows that $\theta=(\lambda_1,\lambda_2,\lambda_3,\lambda_4,\varsigma_{1,1},\varsigma_{1,2},\varsigma_{2,2})$ is globally identified.

The examples just presented reassure the importance of Theorem 2 as a tool for testing the identification status of θ in MCFA-SMN models under *simple structure*. In Example 5 the theorem confirms the obvious identification of θ when the $G=2$ groups enter together in the model, also the test for both groups reduce to the rank test of

Reilly (1995) for only the first group, $g = 1$ (a desired behavior since the parameters in the first group are all identified and the second group, $g = 2$, does not add new parameters to the complete model). Example 6 involves the analysis of an important situation where the parameters of neither groups ($g = 1$ or $g = 2$) are fully identified but complement each other to achieve global identification of θ. It is clear that the normal equations related to group $g = 2$ give the information needed to identify $\varsigma_{1,1}$ and $\varsigma_{2,2}$ in the other group. It is then obvious from the results of Example 4 that the single group $g = 1$ is now in position to identify all of its remaining parameters, consequently leading the parameters of group $g = 2$ to identification. Finally, Theorem 2 provides a easy and computationally simple method that avoids direct manipulation of the normal equations in a search to confirm parameter identification.

Still concerning the applications of Theorem 2, for a single group ($G = 1$), Reilly (1995) brings a thorough discussion of his rank rule directed to partially identified parameters. The author states a second proposition that permits the application of the rank rule to single out identified parameters in the model. Also the author discusses the possibility of correlated parameters in the variance matrix of specific latent factors when all parameters are identified and there exist more normal equation in $\mathscr{P}$ (here considering $G = 1$) than free parameters in the model, *i.e.* the matrix R_1 is of full rank and has more rows than columns. In that case, normal equations can be freed from $\mathscr{P}$, what corresponds to make some $\psi_{ij}^{(g)}$ different from zero in the system (2.9). We shall not discuss those details for the MCFA-SMN model, although we assure they can be successfully applied to the study of this class of models. For more information on this topic, we refer the reader to the original source of Reilly (1995).

2.4 Estimation

As an inherently latent variable model, factor analysis naturally accommodates itself in the framework of the Expectation-Maximization (EM) algorithm (DEMPSTER *et al.*, 1977; RUBIN and THAYER, 1982). The seminal work of Rubin and Thayer (1982) on maximum likelihood estimation of factor analysis models via EM algorithm popularized the method in this field of multivariate analysis, with recent studies still revealing many good properties of their algorithm (ADACHI, 2013; HAYASHI and LIANG, 2014, among others). Variants of Rubin and Thayer (1982)'s EM algorithm have appeared in the literature with several purposes, including strategies to suit different sets of constraints into the parameter matrices of factor analysis models (JAMSHIDIAN and JENNRICH, 1994), to deal with non-normal response variables (MONTANARI and VIROLI, 2010; ZHANG *et al.*, 2014) and also in the scope of multiple factor analysis models (DE VITO, 2016).

For the estimation of CFA models Rubin and Thayer (1982)'s algorithm restricts the covariance of common latent factors to be diagonal or unconstrained, in which cases the M-step of the algorithm has closed form. Jamshidian and Jennrich (1994) propose a modified version of Rubin and Thayer (1982)'s algorithm to accommodate linear restrictions on the covariance of common latent factors, although their solution relies on numerical methods. In the multiple factor analysis models setting, our literature review has not shown any EM algorithm for confirmatory models.

We shall present an Expectation-Conditional-Maximization (ECM) algorithm, using the theory proposed by Meng and Rubin (1993), capable of estimating MCFA-SMN models with a broad range of invariance between parameters in any of the G groups being modeled. The proposed algorithm includes Rubin and Thayer (1982)'s algorithm as a special case when $G = 1$ and the SMN distribution assumed for the MCFA-SMN model coincides with the multivariate normal distribution. The algorithms of Zhang, Li, and Liu (2014) for estimation of their TFA model and of De Vito (2016) for estimation of their multi-study factor model also appears as special cases of our proposed algorithm when viewing the MCFA-SMN as an exploratory factor analysis model.

2.4.1 ECM algorithm

In factor analysis the complete-data structure necessary for the EM algorithm is achieved by treating the matrix of common latent factor as missing data (RUBIN and THAYER, 1982). Although, in comparison to the traditional factor analysis formulation, the MCFA-SMN model have its latent space expanded with the introduction of vectors of independent mixing variables following a common positive univariate distribution. Hence, in the MCFA-SMN setting the complete data results from the joint distribution of (Y_{ig}, Z_{ig}, U_{ig}), where $g \in \{1, \ldots, G\}$ indexes the group where the i-th unit belong. From the stochastic representation given in Definition 1, the joint distribution of (Y_{ig}, Z_{ig}, U_{ig})

can be hierarchically represented as

$$\begin{aligned} Y_{ig}|Z_{ig}=z_{ig},U_{ig}=u_{ig} &\sim \mathsf{N}_{p_g}\left(\mu_g+\Lambda_g z_{ig},\frac{1}{u_{ig}}\Psi_g\right), \\ Z_{ig}|U_{ig}=u_{ig} &\sim \mathsf{N}_{k_g}\left(0,\frac{1}{u_{ig}}\zeta_g\right), \\ U_{ig} &\sim \mathsf{H}(\cdot|\nu),\ g=1,\ldots,G. \end{aligned} \tag{2.13}$$

A complete-data sample from (Y_{ig},Z_{ig},U_{ig}) is denoted by $y_{ig}^{(c)}=(y_{ig}^\top,z_{ig}^\top,u_{ig})^\top$ and the vector of all complete-observations is denoted by $y^{(c)}=(y_{11}^{(c)\top},\ldots,y_{n_1 1}^{(c)\top},\ldots,y_{1G}^{(c)\top}$, $y_{n_G G}^{(c)\top})^\top$. Hence, from representation (2.13), we deduce the complete-data log-likelihood as

$$\ell_c(\theta)=\sum_{g=1}^{G}\left[\ell_{\lambda,\psi}^{(g)}(\theta)+\ell_{\varsigma}^{(g)}(\theta)\right]+\sum_{g=1}^{G}\sum_{i=1}^{n_g}\log\mathsf{H}(u_{ig}|\nu). \tag{2.14}$$

where the last term is a constant, since ν is known by assumption, and the quantities $\ell_{\lambda,\psi}^{(g)}(\theta)$ and $\ell_{\varsigma}^{(g)}(\theta)$ are defined, for $g=1,\ldots,G$, in term of the statistics

$$\boldsymbol{S}_{uy}^{(g)}=\sum_{i=1}^{n_g}u_{ig}y_{ig}y_{ig}^\top,\quad \boldsymbol{S}_{uz}^{(g)}=\sum_{i=1}^{n_g}u_{ig}z_{ig}z_{ig}^\top,\quad \boldsymbol{S}_{uzy}^{(g)}=\sum_{i=1}^{n_g}u_{ig}z_{ig}y_{ig}^\top \tag{2.15}$$

as

$$\ell_{\lambda,\psi}^{(g)}(\theta)=\ \mathsf{tr}\left(\Psi_g^{-1}\Lambda_g\boldsymbol{S}_{uzy}^{(g)}\right)-\frac{1}{2}\mathsf{tr}\left(\Lambda_g^\top\Psi_g^{-1}\Lambda_g\boldsymbol{S}_{uz}^{(g)}\right)-\frac{1}{2}\mathsf{tr}\left(\Psi_g^{-1}\boldsymbol{S}_{uy}^{(g)}\right)-\frac{n_g}{2}\log|\Psi_g| \tag{2.16}$$

and

$$\ell_{\varsigma}^{(g)}(\theta)=-\frac{1}{2}\mathsf{tr}\left(\zeta_g^{-1}\boldsymbol{S}_{uz}^{(g)}\right)-\frac{n_g}{2}\log|\zeta_g|. \tag{2.17}$$

The Q-function is defined as the expectation of the complete-data log-likelihood taken with respect to the conditional distribution of the missing data given the observed data and a previously known vector $\theta^{(k)}$, the update of θ in the $(k-1)$-th iteration of the algorithm (MENG and RUBIN, 1993). Define $\mathscr{E}_{Z,U|Y,\theta^{(k)}}(\cdot)$, the expectation taking with respect to the conditional distribution of the latent variables given the observed data. Hence, the Q-function results as

$$\mathsf{Q}(\theta|\theta^{(k)})=\sum_{g=1}^{G}\left[\hat{\ell}_{\lambda,\psi}^{(g)}(\theta)+\hat{\ell}_{\varsigma}^{(g)}(\theta)\right]+\sum_{g=1}^{G}\sum_{i=1}^{n_g}\log\mathsf{H}(u_{ig}|\nu), \tag{2.18}$$

where $\hat{\ell}^{(g)}_{\lambda,\psi}(\theta) = \mathscr{E}_{Z,U|Y,\theta^{(k)}}\left[\ell^{(g)}_{\lambda,\psi}(\theta)\right]$ and $\hat{\ell}^{(g)}_{\varsigma}(\theta) = \mathscr{E}_{Z,U|Y,\theta^{(k)}}\left[\ell^{(g)}_{\varsigma}(\theta)\right]$.

The Q-function defined in (2.18) uses the conditional distribution of $(Z_{ig}, U_{ig})|Y_{ig} = y_{ig}, \theta = \theta^{(k)}$, $g = 1, \ldots, G$. The expectation taken with respect to the desired joint distribution can be simply obtained using the properties of conditional expectations given by

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} zu\, \mathsf{f}_{Z,U|Y,\theta^{(k)}}(z,u)dudz = \int_{-\infty}^{\infty} z\left[\int_{0}^{\infty} u\, \mathsf{f}_{U|Y,\theta^{(k)}}(u)du\right]\mathsf{f}_{Z|U,Y,\theta^{(k)}}(z)dz. \tag{2.19}$$

The conditional distribution of $Z_{ig}|U_{ig} = u_{ig}, Y_{ig} = y_{ig}, \theta = \theta^{(k)}$, $g = 1, \ldots, G$, could be derived from the known conditional distributions presented in (2.13) together with an application of the Bayes' theorem. Let $\hat{\mathbf{\Lambda}}_g = \mathbf{\Lambda}_g(\theta^{(k)})$, $\hat{\mathbf{\Psi}}_g = \mathbf{\Psi}_g(\theta^{(k)})$ and $\hat{\zeta}_g = \zeta_g(\theta^{(k)})$, then

$$Z_{ig}|Y_{ig} = y_{ig}, U_{ig} = u_{ig}, \theta = \theta^{(k)} \sim \mathsf{N}_{k_g}\left(\boldsymbol{C}_g^{-1}b_{ig}, \frac{1}{u_{ig}}\boldsymbol{C}_g^{-1}\right), \tag{2.20}$$

where

$$\boldsymbol{C}_g^{-1} = \left(\hat{\mathbf{\Lambda}}_g^{\top}\hat{\mathbf{\Psi}}_g^{-1}\hat{\mathbf{\Lambda}}_g + \hat{\zeta}_g^{-1}\right)^{-1} \tag{2.21}$$

and

$$b_{ig} = \hat{\mathbf{\Lambda}}_g^{\top}\hat{\mathbf{\Psi}}_g^{-1}y_{ig}. \tag{2.22}$$

Similarly, the conditional distribution of $U_{ig}|Y_{ig} = y_{ig}, \theta = \theta^{(k)}$, $g = 1, \ldots, G$, stems from the Bayes' theorem, using the fact that $Y_{ig}|U_{ig} = u_{ig}, \theta = \theta^{(k)}$ is normally distributed. In the following, we show the expectation $\mathscr{E}_{U|Y,\theta^{(k)}}(U_{ig})$ that results for the SMN distributions presented in Subsection 1.4.2 and that could be assumed for Y_{ig} in the MCFA-SMN model, $g = 1, \ldots, G,$.

- $Y_{ig} \sim \mathsf{t}_{p_g}(0, \mathbf{\Sigma}_g, \nu)$:

$$\mathscr{E}_{U|Y,\theta^{(k)}}(U_{ig}) = \frac{p_g + \nu}{\nu + \mathsf{d}^2(\theta^{(k)}, y_{ig})}, \tag{2.23}$$

- $Y_{ig} \sim \mathsf{SL}_{p_g}(0, \mathbf{\Sigma}_g, \nu)$:

$$\mathscr{E}_{U|Y,\theta^{(k)}}(U_{ig}) = \frac{2}{\mathsf{d}^2(\theta^{(k)}, y_{ig})}\frac{\Gamma\left(\frac{p_g}{2} + \nu + 1, \mathsf{d}^2(\theta^{(k)}, y_{ig})\right)}{\Gamma\left(\frac{p_g}{2} + \nu, \mathsf{d}^2(\theta^{(k)}, y_{ig})\right)}, \tag{2.24}$$

- $Y_{ig} \sim \mathsf{CN}_{p_g}(0, \boldsymbol{\Sigma}_g, \xi, \gamma)$:

$$\mathscr{E}_{U|Y,\theta^{(k)}}(U_{ig}) = \frac{1-\xi+\xi\gamma^{(p_g/2)+1}\exp\left(\frac{1}{2}(1-\gamma)\mathsf{d}^2(\theta^{(k)}, y_{ig})\right)}{1-\xi+\xi\gamma^{p_g/2}\exp\left(\frac{1}{2}(1-\gamma)\mathsf{d}^2(\theta^{(k)}, y_{ig})\right)}, \tag{2.25}$$

where $\mathsf{d}^2(\theta^{(k)}, y_{ig}) = y_{ig}^\top \boldsymbol{\Sigma}_g^{-1}(\theta^{(k)}) y_{ig}$ and $\Gamma(a,b) = \int_0^b t^{a-1}e^{-t}dt$ is the incomplete gamma function. The results above are also presented by Ferreira, Lachos, and Bolfarine (2016), where it is considered a more general probabilistic context involving skewed SMN distribution.

Yet, another useful result present by Ferreira, Lachos, and Bolfarine (2016) is the probabilistic distribution of the quantity $\mathsf{d}^2(\theta^{(k)}, y_{ig})$ given the distribution of the independently observed random variables Y_{ig}. According to Ferreira, Lachos, and Bolfarine (2016) the following results are valid:

- If $Y_{ig} \sim \mathsf{N}_{p_g}(0, \boldsymbol{\Sigma}_g)$, then

$$\mathsf{d}^2(\theta^{(k)}, y_{ig}) \sim \chi^2_{p_g}, \tag{2.26}$$

- If $Y_{ig} \sim \mathsf{t}_{p_g}(0, \boldsymbol{\Sigma}_g, \nu)$, then

$$\mathsf{d}^2(\theta^{(k)}, y_{ig}) \sim p_g \times \mathsf{F}_{p_g,\nu}, \tag{2.27}$$

- If $Y_{ig} \sim \mathsf{SL}_{p_g}(0, \boldsymbol{\Sigma}_g, \nu)$, then

$$\mathsf{P}\left(\mathsf{d}^2(\theta^{(k)}, y_{ig}) < r\right) = \mathsf{P}\left(\chi^2_{p_g} < r\right) - \frac{2^\nu\Gamma(\nu+p_g/2)}{r^\nu\Gamma(p/2)}\mathsf{P}\left(\chi^2_{2\nu+p_g} < r\right), \tag{2.28}$$

- If $Y_{ig} \sim \mathsf{CN}_{p_g}(0, \boldsymbol{\Sigma}_g, \xi, \gamma)$:

$$\mathsf{P}\left(\mathsf{d}^2(\theta^{(k)}, y_{ig})\right) = \xi\mathsf{P}\left(\chi^2_{p_g} < \gamma r\right) + (1-\xi)\mathsf{P}\left(\chi^2_{p_g} < r\right), \tag{2.29}$$

where χ^2_δ is the chi-square distribution with degree of freedom equals to δ and $\mathsf{F}_{\alpha,\beta}$ is the F distribution with degrees of freedom equal to α and β.

Applying the conditional expectation property (2.19), it can be shown the statistics defined in (2.15) have expectation taken with respect to the joint distribution of

$(Z_{ig}, U_{ig})|Y_{ig} = y_{ig}, \theta = \theta^{(k)}$, $g = 1, \ldots, G$, given by

$$\hat{\boldsymbol{S}}_{uy}^{(g)} = \sum_{i=1}^{n_g} \mathscr{E}_{U|Y,\theta^{(k)}}[U_{ig}] y_{ig} y_{ig}^{\top}, \tag{2.30}$$

$$\hat{\boldsymbol{S}}_{uz}^{(g)} = n_g \boldsymbol{C}_g^{-1} + \sum_{i=1}^{n_g} \left[\mathscr{E}_{U|Y,\theta^{(k)}}[U_{ig}] \boldsymbol{C}_g^{-1} b_{ig} b_{ig}^{\top} \boldsymbol{C}_g^{-1} \right], \tag{2.31}$$

$$\hat{\boldsymbol{S}}_{uzy}^{(g)} = \sum_{i=1}^{n_g} \mathscr{E}_{U|Y,\theta^{(k)}}[U_{ig}] \boldsymbol{C}_g^{-1} b_{ig} y_{ig}^{\top}, \tag{2.32}$$

where $\boldsymbol{C}_g^{-1}$ and b_{ig} are defined in (2.21) and (2.22).

Hence, the expected values that enter in the definition of the Q-function (2.18) are simply given by

$$\hat{\ell}_{\lambda,\psi}^{(g)}(\theta) = \text{ tr}\left(\boldsymbol{\Psi}_g^{-1} \boldsymbol{\Lambda}_g \hat{\boldsymbol{S}}_{uzy}^{(g)}\right) - \frac{1}{2}\text{tr}\left(\boldsymbol{\Lambda}_g^{\top} \boldsymbol{\Psi}_g^{-1} \boldsymbol{\Lambda}_g \hat{\boldsymbol{S}}_{uz}^{(g)}\right) - \frac{1}{2}\text{tr}\left(\boldsymbol{\Psi}_g^{-1} \hat{\boldsymbol{S}}_{uy}^{(g)}\right) - \frac{n_g}{2} \log|\boldsymbol{\Psi}_g| \tag{2.33}$$

and

$$\hat{\ell}_{\varsigma}^{(g)}(\theta) = -\frac{1}{2}\text{tr}\left(\boldsymbol{\zeta}_g^{-1} \hat{\boldsymbol{S}}_{uz}^{(g)}\right) - \frac{n_g}{2} \log|\boldsymbol{\zeta}_g|. \tag{2.34}$$

The conditional maximization step (CM-step) of the ECM algorithm operates on a set $\mathscr{G} = \{\mathsf{g}_s(\theta) : s = 1, \ldots, S\}$ of functions $\mathsf{g}_s(\cdot)$ that constrain the parameter of interest for estimation, θ. The functions defined in $\mathscr{G}$ partition θ in S parts: $\theta^{(1)}, \ldots, \theta^{(S)}$. That way, in the $(k-1)$-th iteration of the ECM algorithm the aim of the CM-step is to find a value of $\theta_k^{(s)}$, $s \in \{1, \ldots, S\}$, such that

$$\mathsf{Q}(\theta_k^{(s)}|\theta_{k-1}) \geq \mathsf{Q}(\theta|\theta_{k-1}) \;\text{ with }\; \theta \in \{\theta \in \Theta : \mathsf{g}_s(\theta) = \mathsf{g}_s(\theta^{(-s)})\}, \tag{2.35}$$

where $\theta^{(-s)}$ is the vector θ with the elements corresponding to its part $\theta^{(s)}$ omitted.

The ECM algorithm will converge to the unconstrained maximum of the log-likelihood (2.5) only if a property of $\mathscr{G}$, called by Meng and Rubin (1993) as the *space filling* property, is verified. Consider $\mathsf{g}_s(\cdot)$ differentiable and with gradient $\nabla \mathsf{g}_s(\theta)$ having full rank at θ in the interior of Θ, $s = 1, \ldots, S$. Then, according to Meng and Rubin (1993) the set $\mathscr{G}$ will be *space filling* if

$$\mathsf{J}(\theta) = \bigcap_{s=1}^{S} \mathsf{J}_s(\theta) = \{0\}, \tag{2.36}$$

where $\mathrm{J}_s(\theta) = \{\nabla \mathrm{g}_s(\theta)\lambda : \lambda \in \mathbb{R}^{d_s}\}$ is the column space of the gradient of $g_s(\theta)$ and d_s is the dimentionality of $g_s(\theta)$. If $\mathscr{G}$ is *space filling*, then it is guaranteed that at any iteration the ECM algorithm will be free to search the parameter space in any direction (MENG and RUBIN, 1993). In the following, we elaborate a set of constraints $\mathscr{G}$ over θ that is *space filling*.

Let a finite sequence $(v_{\jmath})_{\jmath \in \mathscr{S}}$ be called a partition of a vector v if $v = (v_{\jmath})_{\jmath \in \mathscr{S}}$ and the length of v equals the length of $(v_{\jmath})_{\jmath \in \mathscr{S}}$. We shall partition θ in two levels. The first level is $\theta = (\theta_\lambda, \theta_\psi, \theta_\varsigma)$, where θ_λ, θ_ψ and θ_ς are such that for any $g \in \{1, \ldots, G\}$ $\boldsymbol{\Lambda}_g(\theta) = \boldsymbol{\Lambda}_g(\theta_\lambda)$, $\boldsymbol{\Psi}_g(\theta) = \boldsymbol{\Psi}_g(\theta_\psi)$ and $\boldsymbol{\zeta}_g(\theta) = \boldsymbol{\zeta}_g(\theta_\varsigma)$. Since θ has no repeated elements, θ_λ, θ_ψ and θ_ς does not share any of its elements between them. The second level of the partition operates over θ_λ, θ_ψ and θ_ς and is generically defined as

$$\theta_\alpha = \left(\theta_{\alpha(\hbar)}\right)_{\hbar \in \mathscr{H}_\alpha}, \tag{2.37}$$

where α is an index with possible labels λ, ψ or ς and $\mathscr{H}_\alpha$ is a family set over $\{1, \ldots, G\}$ with its elements being possibly sets with at least one element, but the elements of $\mathscr{H}_\alpha$ must not have itself the empty set as an element. In (2.37), if $\hbar \in \mathscr{H}_\alpha$ is given, the element $\theta_{\alpha(\hbar)}$ is a vector comprised by the parameters in θ_α that are shared by groups indexed in $\hbar$. We ought to define $\mathscr{H}_\alpha$ so that it avoids inconsistencies in the CM-steps of the ECM algorithm. For that being so, $\mathscr{H}_\alpha$ must have elements that are subset of groups' index with no intersection between subsets, that guarantees $\left(\theta_{\alpha(\hbar)}\right)_{\hbar \in \mathscr{H}_\alpha}$ is in fact a partition of θ_α. Finally, the elements of $\mathscr{H}_\alpha$ must not have the empty set as an element in order to avoid the meaningless situation where $\theta_{\alpha(\hbar_\varnothing)}$ occurs, with $\varnothing \in \hbar_\varnothing$ and $\hbar_\varnothing \in \mathscr{H}_\alpha$.

Let θ be a p-dimensional vector of parameters partitioned as in (2.37) and $\theta_{(\hbar_\alpha)}$, of length $r \leq p$, $\hbar_\alpha \in \mathscr{H}_\alpha$, be one of its parts. The vector θ can be explicitly expressed in terms of $\theta_{(\hbar_\alpha)}$ as

$$\theta = \boldsymbol{P}_{(\hbar_\alpha)}^\top \boldsymbol{Q}_{(\hbar_\alpha)} \theta_{(\hbar_\alpha)} + \boldsymbol{P}_{(\hbar_\alpha)}^\top \boldsymbol{W}_{(\hbar_\alpha)} \theta_{(-\hbar_\alpha)}. \tag{2.38}$$

where the special matrices $\boldsymbol{P}_{(\hbar_\alpha)}$, $\boldsymbol{Q}_{(\hbar_\alpha)}$ and $\boldsymbol{W}_{(\hbar_\alpha)}$ have the following properties

$$\boldsymbol{P}_{(\hbar_\alpha)}\theta = \begin{bmatrix} \theta_{(\hbar_\alpha)} \\ \theta_{(-\hbar_\alpha)} \end{bmatrix}, \; \boldsymbol{Q}_{(\hbar_\alpha)} = \begin{bmatrix} \boldsymbol{I}_r \\ \mathbf{0}_{(p-r)\times r} \end{bmatrix}, \; \boldsymbol{W}_{(\hbar_\alpha)} = \begin{bmatrix} \mathbf{0}_{r\times(p-r)} \\ \boldsymbol{I}_{p-r} \end{bmatrix}, \tag{2.39}$$

and $\theta_{(-\hbar_\alpha)}$ corresponds to some known reordering of θ devoid of its sub-vector $\theta_{(\hbar_\alpha)}$.

From Equation (2.38) it is immediate that the partial derivative of θ with respect to $\theta_{(\hbar_\alpha)}$, for some $\hbar_\alpha \in \mathscr{H}_\alpha$, is given by

$$\frac{\partial\theta}{\partial\theta^\top_{(\hbar_\alpha)}} = \boldsymbol{P}^\top_{(\hbar_\alpha)}\boldsymbol{Q}_{(\hbar_\alpha)}. \tag{2.40}$$

We now state a proposition with the aim of defining a set $\mathscr{G}$ of constraints over θ and proving that it is *space filling*.

Proposition 2. *The set* $\mathscr{G} = \{g_{(\hbar_\alpha)}(\theta) = \theta_{(\hbar_\alpha)} | \alpha \in \{\lambda, \psi, \varsigma\}, \hbar \in \mathscr{H}_\alpha\}$ *is space filling.*

Proof. Let α be fixed and $\hbar_\alpha \in \mathscr{H}_\alpha$. From Equation (2.38) it can be seen that

$$\mathrm{g}_{(\hbar_\alpha)}(\theta) = \boldsymbol{Q}^\top_{(\hbar_\alpha)}\boldsymbol{P}_{(\hbar_\alpha)}\theta \;\text{ and }\; \mathrm{g}_{(-\hbar_\alpha)}(\theta) = \boldsymbol{W}^\top_{(\hbar_\alpha)}\boldsymbol{P}_{(\hbar_\alpha)}\theta.$$

Hence, their gradients are

$$\nabla\mathrm{g}_{(\hbar_\alpha)}(\theta) = \boldsymbol{Q}^\top_{(\hbar_\alpha)}\boldsymbol{P}_{(\hbar_\alpha)} \;\text{ and }\; \nabla\mathrm{g}_{(-\hbar_\alpha)}(\theta) = \boldsymbol{W}^\top_{(\hbar_\alpha)}\boldsymbol{P}_{(\hbar_\alpha)}.$$

The column spaces of $\nabla\mathrm{g}_{(\hbar_\alpha)}(\theta)$ and $\nabla\mathrm{g}_{(-\hbar_\alpha)}(\theta)$ are certainly orthogonal complements. Hence, any subspace of the column space of $\nabla\mathrm{g}_{(-\hbar_\alpha)}(\theta)$ will also be orthogonal to $\nabla\mathrm{g}_{(\hbar_\alpha)}(\theta)$. Repeating that same process for all parts of θ shows that $\mathscr{G}$ fulfills the sufficient conditions for being *space filling*, as defined by Meng and Rubin (1993). □

Before presenting the ECM algorithm, we give the necessary derivatives for stating the CM-steps. Consider $\hbar_\lambda \in \mathscr{H}_\lambda$, $\hbar_\psi \in \mathscr{H}_\psi$ and $\hbar_\varsigma \in \mathscr{H}_\varsigma$, the desired partial derivatives of the Q-function are the following

$$\frac{\partial\mathrm{Q}(\theta|\theta^{(k)})}{\partial\theta^\top_{(\hbar_\lambda)}} = \sum_{g\in\hbar_\lambda} \frac{\partial\mathrm{Q}(\theta|\theta^{(k)})}{\partial\mathrm{vec}(\boldsymbol{\Lambda}_g)^\top}\frac{\partial\mathrm{vec}(\boldsymbol{\Lambda}_g)}{\partial\theta^\top_{(\hbar_\lambda)}} \tag{2.41}$$

$$\frac{\partial \mathrm{Q}(\theta|\theta^{(k)})}{\partial \theta^{\top}_{(\hbar_\psi)}} = \sum_{g\in\hbar_\psi} \frac{\partial \mathrm{Q}(\theta|\theta^{(k)})}{\partial \mathrm{vec}(\Psi_g)^{\top}} \frac{\partial \mathrm{vec}(\Psi_g)}{\partial \mathrm{diag}(\Psi_g)^{\top}} \frac{\partial \mathrm{diag}(\Psi_g)}{\partial \theta^{\top}_{(\hbar_\psi)}} \tag{2.42}$$

$$\frac{\partial \mathrm{Q}(\theta|\theta^{(k)})}{\partial \theta^{\top}_{(\hbar_\varsigma)}} = \sum_{g\in\hbar_\varsigma} \frac{\partial \mathrm{Q}(\theta|\theta^{(k)})}{\partial \mathrm{vec}(\zeta_g)^{\top}} \frac{\partial \mathrm{vec}(\zeta_g)}{\partial \mathrm{vech}(\zeta_g)^{\top}} \frac{\partial \mathrm{vech}(\zeta_g)}{\partial \theta^{\top}_{(\hbar_\varsigma)}}. \tag{2.43}$$

The further development of the derivatives in (2.41), (2.42) and (2.43) depends only on well known results of matrix calculus, which could be found in Magnus and Neudecker (1985) and Magnus (2010). The final results are

$$\begin{aligned}
\frac{\partial \mathrm{Q}(\theta|\theta^{(k)})}{\partial \theta^{\top}_{(\hbar_\lambda)}} = & \sum_{g\in\hbar_\lambda} \mathrm{vec}\left(\Psi_g^{-1}\hat{S}^{(g)}_{uzy}\right)^{\top} P^{(g)\top}_{(\hbar_\lambda)} Q^{(g)}_{(\hbar_\lambda)} \\
& - \theta^{\top}_{(\hbar_\lambda)} \sum_{g\in\hbar_\lambda} Q^{(g)\top}_{(\hbar_\lambda)} P^{(g)}_{(\hbar_\lambda)} \left(\hat{S}^{(g)}_{uz} \otimes \Psi_g^{-1}\right) P^{(g)\top}_{(\hbar_\lambda)} Q^{(g)}_{(\hbar_\lambda)} \\
& - \sum_{g\in\hbar_\lambda} \mathrm{vec}\left(\Lambda_g\right)^{\top}_{(-\hbar_\lambda)} W^{(g)\top}_{(\hbar_\lambda)} P^{(g)}_{(\hbar_\lambda)} \left(\hat{S}^{(g)}_{uz} \otimes \Psi_g^{-1}\right) P^{(g)\top}_{(\hbar_\lambda)} Q^{(g)}_{(\hbar_\lambda)},
\end{aligned} \tag{2.44}$$

$$\begin{aligned}
\frac{\partial \mathrm{Q}(\theta|\theta^{(k)})}{\partial \theta^{\top}_{(\hbar_\psi)}} = & \frac{1}{2} \sum_{g\in\hbar_\psi} \left[\mathrm{vec}\left(2\Lambda_g\hat{S}^{(g)}_{uzy} - \Lambda_g\hat{S}^{(g)}_{uz}\Lambda_g^{\top} - \hat{S}^{(g)}_{uy}\right)^{\top} + n_g\mathrm{vec}(\Psi_g)^{\top}\right] \\
& \times \left(\Psi_g^{-1} \otimes \Psi_g^{-1}\right) B_{p_g} P^{(g)\top}_{(\hbar_\psi)} Q^{(g)}_{(\hbar_\psi)},
\end{aligned} \tag{2.45}$$

$$\frac{\partial \mathrm{Q}(\theta|\theta^{(k)})}{\partial \theta^{\top}_{(\hbar_\varsigma)}} = -\frac{1}{2} \sum_{g\in\hbar_\varsigma} \left[n_g\mathrm{vec}(\zeta_g)^{\top} - \mathrm{vec}\left(\hat{S}^{(g)}_{uz}\right)^{\top}\right] \left(\zeta_g^{-1} \otimes \zeta_g^{-1}\right) D_{k_g} P^{(g)\top}_{(\hbar_\varsigma)} Q^{(g)}_{(\hbar_\varsigma)}, \tag{2.46}$$

where the matrices B_{p_g}, of order ${p_g}^2 \times p_g$, and D_{k_g}, of order ${k_g}^2 \times k_g(k_g+1)/2$, are the diag matrix and duplication matrix, respectively, defined in Appendix A.

The ECM algorithm for the estimation of MCFA-SMN models will restrict attention to those models where θ is constrained according to the functions composing the set $\mathscr{G}$ defined in Proposition 2. That being so, consider an initial value $\theta^{(0)}$ in the parameter space, the k-th iteration of the algorithm is defined as follows

- **Initialization**: Based on the results of Adachi (2013), consider as initial values for the parameters associated with each group $g \in \{1,\ldots,G\}$ the estimates of an orthogonal factor analysis model. If the resultant estimates are proper, use them to initialize the matrices $\Lambda_g(\theta^{(0)})$, $\Psi_g(\theta^{(0)})$ and $\zeta_g(\theta^{(0)})$, with its fixed parameters substituted by their respective pre-assigned values;

- **E-step**: At the k-th iteration, compute the Q-function using its definition (2.18) together with Equations (2.33) and (2.34);

- **CM-step**: Let $\hbar_\lambda \in \mathscr{H}_\lambda$, $\hbar_\psi \in \mathscr{H}_\psi$ and $\hbar_\varsigma \in \mathscr{H}_\varsigma$. Update each part of $\theta^{(k)}$ using the expressions

$$\begin{aligned}\theta_{(\hbar_\lambda)}^{(k+1)^\top} = \sum_{g\in\hbar_\lambda} &\left\{\left[\text{vec}\left(\hat{\boldsymbol{\Psi}}_g^{-1}\hat{\boldsymbol{S}}_{uzy}^{(g)}\right)^\top - \text{vec}\left(\hat{\boldsymbol{\Lambda}}_g\right)_{(-\hbar_\lambda)}^\top \boldsymbol{W}_{(\hbar_\lambda)}^{(g)\top}\boldsymbol{P}_{(\hbar_\lambda)}^{(g)}\left(\hat{\boldsymbol{S}}_{uz}^{(g)}\otimes\hat{\boldsymbol{\Psi}}_g^{-1}\right)\right]\right.\\ &\left.\times\, \boldsymbol{P}_{(\hbar_\lambda)}^{(g)\top}\boldsymbol{Q}_{(\hbar_\lambda)}^{(g)}\right\}\left[\sum_{g\in\hbar_\lambda}\boldsymbol{Q}_{(\hbar_\lambda)}^{(g)\top}\boldsymbol{P}_{(\hbar_\lambda)}^{(g)}\left(\hat{\boldsymbol{S}}_{uz}^{(g)}\otimes\hat{\boldsymbol{\Psi}}_g^{-1}\right)\boldsymbol{P}_{(\hbar_\lambda)}^{(g)\top}\boldsymbol{Q}_{(\hbar_\lambda)}^{(g)}\right]^{-1},\end{aligned} \tag{2.47}$$

$$\theta_{(\hbar_\psi)}^{(k+1)^\top} = -\left(\sum_{g\in\hbar_\psi}\frac{1}{n_g}\right)\sum_{g\in\hbar_\psi}\text{vec}\left[\left(2\hat{\boldsymbol{\Lambda}}_g\hat{\boldsymbol{S}}_{uzy}^{(g)} - \hat{\boldsymbol{\Lambda}}_g\hat{\boldsymbol{S}}_{uz}^{(g)}\hat{\boldsymbol{\Lambda}}_g^\top - \hat{\boldsymbol{S}}_{uy}^{(g)}\right)^\top\right]\boldsymbol{P}_{(\hbar_\psi)}^{(g)\top}\boldsymbol{Q}_{(\hbar_\psi)}^{(g)}, \tag{2.48}$$

$$\theta_{(\hbar_\varsigma)}^{(k+1)^\top} = \max_{\theta_{(\hbar_\varsigma)}}\left\{-\frac{1}{2}\sum_{g\in\hbar_\varsigma}\text{tr}\left(\boldsymbol{\zeta}_g^{-1}\boldsymbol{S}_{uz}^{(g)}\right) + n_g\log|\boldsymbol{\zeta}_g^{-1}|\right\}. \tag{2.49}$$

- **Stopping rule**: Stop the algorithm if $\sqrt{\sum_{i=1}^r(\theta_i^{(k+1)} - \theta_i^{(k)})^2} < \varepsilon$, for some small positive constant ε, usually $\varepsilon = 10^{-6}$.

In the CM-step it remains to prove the existence of the matrix inverse appearing in the first update Equation (2.47). The next proposition will serve to this task.

Proposition 3. *For $\hbar_\lambda \in \mathscr{H}_\lambda$, the matrix $\sum_{g\in\hbar_\lambda}\boldsymbol{Q}_{(\hbar_\lambda)}^{(g)\top}\boldsymbol{P}_{(\hbar_\lambda)}^{(g)}\left(\hat{\boldsymbol{S}}_{uz}^{(g)}\otimes\boldsymbol{\Psi}_g^{-1}\right)\boldsymbol{P}_{(\hbar_\lambda)}^{(g)\top}\boldsymbol{Q}_{(\hbar_\lambda)}^{(g)}$ is non-singular.*

Proof. Let g be fixed in $\{1,\ldots,G\}$. Lemma 1 of Adachi (2013) proves that $\hat{\boldsymbol{S}}_{uz}^{(g)}$ is positive definite. Since $\boldsymbol{\Psi}_g^{-1}$ is guaranteed to be proper, then $\hat{\boldsymbol{S}}_{uz}^{(g)}\otimes\boldsymbol{\Psi}_g^{-1}$ is positive definite. Consequently, $\boldsymbol{Q}_{(\hbar_\lambda)}^{(g)\top}\boldsymbol{P}_{(\hbar_\lambda)}^{(g)}\left(\hat{\boldsymbol{S}}_{uz}^{(g)}\otimes\boldsymbol{\Psi}_g^{-1}\right)\boldsymbol{P}_{(\hbar_\lambda)}^{(g)\top}\boldsymbol{Q}_{(\hbar_\lambda)}^{(g)} > 0$. Since it occurs for any g, the validity of the proposed result is verified. □

The starting values for the ECM algorithm are motivated by a set of theorems proposed by Adachi (2013), which states that if in a CFA model $(G = 1)$ the matrices $\hat{\boldsymbol{S}}_{uy}^{(1)}$, $\boldsymbol{\Psi}_1(\theta^0)$ and $\boldsymbol{\zeta}(\theta^0)$ are positive definite, at convergence the EM algorithm of Rubin and Thayer (1982) will generate only proper solutions. That is to say, there would not be negative variances in $\boldsymbol{\Psi}_1(\hat{\theta})$, known in the literature as Heywood cases (HAYASHI and LIANG, 2014), and $\boldsymbol{\zeta}(\hat{\theta})$ would be positive-definite. When $\hat{\boldsymbol{S}}_{uy}^{(1)}$ is non-negative definite the EM algorithm of Rubin and Thayer (1982) could converge to a $\hat{\theta}$ such that some diagonal elements of $\boldsymbol{\Psi}_1(\hat{\theta})$ equal zero, although being never negative. The theorems in Adachi (2013) are proved only for the case $G = 1$. Although, it is intuitive that it would work for MCFA-SMN models, with the conditions on the theorems of Adachi (2013) extended to positive definiteness or non-negative definiteness of $\hat{\boldsymbol{S}}_{uy}^{(g)}$, $g = 1, \ldots, G$.

2.5 Standard errors

In this section we shall present two methods for estimation of standard errors of $\hat{\theta}$ in the MCFA-SMN model. The first method uses the empirical Fisher information matrix, as proposed by Meilijson (1989) in the context of the EM algorithm. The second method is a numerical approximation of the observed information matrix called *central difference approximation* used in factor analysis by Jamshidian (1997).

Let $\mathsf{H}(\theta)$ be the Fisher information of θ. Meilijson (1989) proposed to use the empirical Fisher information, denoted by $\hat{\mathsf{H}}(\theta)$, as a consistent estimate of $\mathsf{H}(\theta)$. Based on a MCFA-SMN model, $\hat{\mathsf{H}}(\theta)$ is given by

$$\hat{\mathsf{H}}(\theta) = \sum_{g=1}^{G}\sum_{i=1}^{n_g} s(y_{ig}|\theta)s(y_{ig}|\theta)^\top + \sum_{g=1}^{G}\frac{1}{n_g}S(y_g|\theta)S(y_g|\theta)^\top, \tag{2.50}$$

where y_g is the vector of observation for the g-th group, $s(y_{ig}|\theta)$ is the score of the i-th individual in the g-th group and $S(y_g|\theta) = \sum_{i=1}^{n_g} s(y_{ig}|\theta)$. Observe that differently from the Fisher Information, $\mathsf{H}(\theta)$, the empirical Fisher information, $\hat{\mathsf{H}}(\theta)$, does not involve second order derivatives.

Additionally, considering the maximum likelihood estimate $\hat{\theta}$ we have

$$\hat{\mathsf{H}}(\hat{\theta}) = \sum_{g=1}^{G}\sum_{i=1}^{n_g} s(y_{ig}|\hat{\theta})s(y_{ig}|\hat{\theta})^\top, \tag{2.51}$$

and the standard errors of $\hat{\theta}$ will be approximated by the square root of the diagonal of $\hat{\mathsf{H}}^{-1}(\hat{\theta})$.

Using the results of Meilijson (1989) on the properties of the EM algorithm, we have that, for θ_0 in the parametric space Θ, the following relation holds

$$\frac{\partial}{\partial\theta}\mathsf{Q}(\theta|\theta_0)\Big|_{\theta=\theta_0} = \sum_{g=1}^{G} S(y|\theta_0). \tag{2.52}$$

A comment made by Meilijson (1989), and more widely discussed by Jamshidian (1997), is that $\mathsf{H}(\theta)$ can be numerically approximated using a process involving successive evaluations of the score function (2.52) at perturbations of the estimate $\hat{\theta}$. Let

$\tilde{\theta}$ be the estimate $\hat{\theta}$ with its k-th element perturbed by a small amount ε. According to Meilijson (1989) the expression

$$\frac{1}{\varepsilon}\sum_{g=1}^{G}\left[S(y|\tilde{\theta})-S(y|\hat{\theta})\right], \tag{2.53}$$

will approximate the k-th column of $\mathrm{H}(\theta)$.

Jamshidian (1997) restated the result (2.53) in the following way: define the column vector $d_j(\theta)$ as

$$d_j(\theta)=\frac{\mathrm{g}(\theta+\varepsilon_j e_j)-\mathrm{g}(\theta-\varepsilon_j e_j)}{2\varepsilon_j},\ j=1,\ldots,q, \tag{2.54}$$

where $\mathrm{g}(\theta)$ is the gradient of the log-likelihood $\ell(\theta)$ evaluated at θ, e_j is a vector with all its entries equal to zero except for the j-th entry, which is equal to one, ε_j is a small number and q is the dimension of θ. Hence, according to Jamshidian (1997) the Fisher information matrix $\mathrm{H}(\theta)$ is approximated by

$$\tilde{\mathrm{H}}(\theta)=\frac{\boldsymbol{D}(\theta)+\boldsymbol{D}^{\top}(\theta)}{2}, \tag{2.55}$$

where $\boldsymbol{D}(\theta)$ is the $q\times q$ matrix with columns equal to $d_j(\theta)$. As mentioned by Lin *et al.* (2014), $\boldsymbol{D}(\theta)$ can be used to approximate $\mathrm{H}(\theta)$, although there are situations where $\boldsymbol{D}(\theta)$ could result in a non-symmetrical matrix. Hence, (2.55) is preferable.

Meilijson (1989) stated that the choice of value for ε_j should be guided by the rule of thumbs of the differential calculus. Jamshidian (1997) and Lin *et al.* (2014) suggested to use $\varepsilon_j=\max(\eta,\eta|\theta_j|)$, with $\eta=10^{-4}$.

Equations (2.50) and (2.55) make use of relation (2.52) to approximate $\mathrm{H}(\theta)$. Hence the derivative of the Q-function, presented in Equations (2.41), (2.42) and (2.43), can be readily used to obtain approximations of the standard errors of $\hat{\theta}$.

3 Simulation

3.1 Resumo da seção

Nesta sessão apresentamos um estudo de simulação para verificar as propriedades dos estimadores dos parâmetros do modelo MCFA-SMN sob amostras finitas. O estudo de simulação foca nos modelos MCFA-N, MCFA-t, MCFA-CN e MCF-SL, com fatores latentes seguindo distribuição normal, t-*Student*, normal contaminada e *slash*, respectivamente. Além disso, nosso estudo de simulação também visa avaliar a performance dos erros padrão obtidos de acordo com os dois métodos apresentados na Subsessão 2.5.

3.2 Simulation design

In this section we shall present results of Monte Carlo simulation studies designed to evaluate the finite sample performance of parameter estimates obtained through the estimators developed in Section 2.4. All simulations were developed with the R software 3.5.0 (R CORE TEAM, 2018). The codes used in the simulation are available by request to the authors.

We use a real data set to guide the model structure to be simulated and also to setup the true values for the parameters in the model. Additionally, we discuss a kind of factor indeterminacy that is an issue of interest in simulation of FA models, namely, the label switch and sign change of common latent factors.

3.2.1 Data set

In order to simulate interpretable FA models we based our simulation study on a real data set commonly appearing in statistical papers on the subject, as for example: Meredith (1964), Jöreskog (1971), Sörbom (1974), Zhong and Yuan (2011) and Lai and Zhang (2017). To motivate the choice of parameter set and factor structure used in the simulations, we first give a brief description of the data set used, which was originally described by Holzinger and Swineford (1939).

Holzinger and Swineford (1939) collected data on 301 students enrolled in two schools, Pasteur $(n = 156)$ and Grant-White $(n = 145)$. The students were from different socio-economic status, with the Pasteur School enrolling student from families with low income and the Grant-White enrolling students from families of middle class. A test comprising 25 items was administered to each one of the 301 students, with the aim of measuring 5 latent factors.

Based on Holzinger and Swineford (1939)'s study, Jöreskog (1971) selected 9 items considered indicators of three common latent factors, interpreted as space, verbal and memory factors. Jöreskog (1971) advocated that a MCFA model for two groups and with invariant Λ_g, ζ_g and Ψ_g, $g = 1,2$, would fit well the data. Additionally, the author proposed a *simple structure* for the loading matrix and set the metric of observed variables by fixing the first loading of each column of the loading matrix equal to one. That way, the model matrices were defined as

$$\Lambda_g = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{2,1} & 0 & 0 \\ \lambda_{3,1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{5,2} & 0 \\ 0 & \lambda_{6,2} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{8,3} \\ 0 & 0 & \lambda_{9,3} \end{bmatrix}, \ \zeta_g = \begin{bmatrix} \varsigma_{1,1} & & \\ \varsigma_{2,1} & \varsigma_{2,2} & \\ \varsigma_{3,1} & \varsigma_{3,2} & \varsigma_{3,3} \end{bmatrix}, \ \Psi_g = \text{diag}(\psi_{j,j})_{j=1}^{9}, \tag{3.1}$$

where $g = 1,2$.

For $g = 1,2$, consider $k_g = 3$ the number of latent factors and $p_g = 9$ the number of observed variables entering in the specification of the CFA model related to the g-th group.

The identification of parameters in the MCFA model defined by Equation (3.1) can be proved by means of Theorem 2. This theorem guarantees that the model's parameters are identified when the Jacobian matrix R_2, defined in Equation (2.10), is of full column rank. For the aimed model, the matrix R_2 is of dimension 72×15 and has rank equals to 15, hence R_2 is of full column rank and, by Theorem 2, the MCFA model defined in (3.1) has all of its parameters identified.

We estimated the parameters of the model proposed by Jöreskog (1971) using the functionalities of the R package *lavaan* (ROSSEEL, 2012), which includes a function for maximum likelihood estimation of MCFA models and also brings Holzinger and

Swineford (1939)' data set. The point estimates we obtained for the parameters are given in Table 1.

Table 1: Parameter estimates for the MFCA model of Jöreskog (1971).

$\boldsymbol{\Lambda}$		$\boldsymbol{\zeta}$		$\boldsymbol{\Psi}$	
Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
$\lambda_{2,1}$	0.6048	$\varsigma_{1,1}$	0.5465	$\psi_{1,1}$	0.4469
$\lambda_{3,1}$	0.8455	$\varsigma_{2,1}$	0.3084	$\psi_{2,2}$	0.7935
$\lambda_{5,2}$	1.0060	$\varsigma_{3,1}$	0.1968	$\psi_{3,3}$	0.6027
$\lambda_{6,2}$	0.9873	$\varsigma_{2,2}$	0.7033	$\psi_{4,4}$	0.2901
$\lambda_{8,3}$	1.2306	$\varsigma_{3,2}$	0.1670	$\psi_{5,5}$	0.2816
$\lambda_{9,3}$	1.1066	$\varsigma_{3,3}$	0.3439	$\psi_{6,6}$	0.3079
				$\psi_{7,7}$	0.6497
				$\psi_{8,8}$	0.4725
				$\psi_{9,9}$	0.5722

3.2.2 Scenarios for simulation

The models we assumed in our simulation study follow the same latent structure of the model proposed by Jöreskog (1971) for analyzing Holzinger and Swineford (1939)' data set. That is to say, a simultaneous factor analysis for $G = 2$ groups with invariant model matrices structured as in (3.1). We considered four different scenarios for simulation, which are described below:

- MCFA-N: Latent factors following a multivariate normal distribution;
- MCFA-t: Latent factors following a multivariate t-Student distribution with $\nu = 4$;
- MCFA-CN: Latent factors following a multivariate contaminated normal distribution with $\xi = 0.5$ and $\gamma = 0.5$;
- MCFA-SL: Latent factors following a multivariate slash distribution with $\nu = 4$.

In all the above scenarios, the choices of values for the parameters indexing the distribution of the mixing variable (that is to say, the choices of ν, ξ and γ) was made to obtain SMN distributions with considerably heavier tails than the normal distribution. Each model in the scenarios above has 21 parameter with true values set equal to the estimates obtained for the model of Jöreskog (1971) and displayed in Table 1. For each scenario we generated artificial samples with sizes fixed at $n = 200, 400, 600, 800$ and 1000. The number of replications for each sample size was

$R = 5000$. The samples for the MCFA-SMN models were generated using the hierarchical structure

$$\begin{aligned} Y_{ig}|Z_{ig} = z_{ig}, U_{ig} = u_{ig} &\sim \mathsf{N}_{p_g}\left(\mu_g + \Lambda_g z_{ig}, \frac{1}{u_{ig}}\Psi_g\right), \\ Z_{ig}|U_{ig} = u_{ig} &\sim \mathsf{N}_{k_g}\left(0, \frac{1}{u_{ig}}\zeta_g\right), \\ U_{ig} &\sim \mathsf{H}(\cdot|\nu),\ g = 1,2, \end{aligned} \tag{3.2}$$

with the multivariate normal samples being generate using the R package *mvtnorm* (some description of the package can be obtained in Amatya and Demirtas (2015)).

The initial values for the parameters in Λ_g, Ψ_g and ζ_g, $g = 1,2$ were taken as the estimates of parameters in two separate exploratory factor analysis under normality, hence the ζ_g were taken as the identity matrix of order 3×3, $g = 1,2$. The ECM algorithm stopped when the k-th and $(k+1)$-th update of θ were such that the euclidean distance $\sqrt{\sum_{i=1}^{21}(\theta_i^{(k+1)} - \theta_i^{(k)})^2}$ was less than 10^{-6}, where θ_i is the i-th entry of θ.

3.2.3 Factor indeterminacy

Here we call attention for an important matter in simulation of FA models, namely, the indeterminacy of common latent factors. This problem is not an issue in our scenarios of simulation, since our model have fully identified parameters and the rotational indeterminacy of the factor loading matrices is resolved with the identification restriction we imposed for the simulated models (PEETERS, 2012).

The uniqueness problem presented in Definition 2 posits an issue for simulation studies of CFA models. According to Definition 2, there exist equivalent solutions to the parameters in the loading matrix and common latent factors' covariance matrix that differ only by an orthogonal rotation or sign change. Those sources of factor indeterminacy were studied by Myers *et al.* (2016).

Myers *et al.* (2016) studied that problem of factor indeterminacy in the context of simulation and proposed to choose the solution with smallest Mean Square Error (MSE). Since in simulation studies the true values of parameters are known, its is possible to list all equivalent solutions that arise by changing signs or reordering the rows and columns of common latent factor's covariance matrix and then calculate their associated MSE. Myers *et al.* (2016) developed an R package called *REREFACT* that execute this task.

3.3 Results

In this section we present the results of simulation for the four MCFA-SMN selected models, namely, the MCFA-N, MCFA-t, MCF-CN and MCFA-SL models. The simulation focus in the analysis of bias, Mean Square Error (MSE) and Monte Carlo standard errors (MC SE) of the proposed estimators as well as in the performance of standard errors and confidence intervals obtained by means of the methods described in Subsection 2.5, namely, the Central Difference Method (CDM) and the Empirical Fisher Information (EFI).

In order to simplify the description of the simulation results we shall adopt the following nomenclature: MSE = mean square error, MC SE = Monte Carlo standard error, CDM SE = average standard error obtained using the central difference method, EFI SE = average standard error obtained using the empirical Fisher information matrix, Prob. CDM = coverage probability of 95% confidence interval (CI) constructed through the central difference method, Prob. EFI = coverage probability of 95% CI obtained through the empirical Fisher information matrix.

3.3.1 Finite sample properties

In order to evaluate the finite sample properties of the estimators developed in Section 2.4 we calculated the bias and MSE of the estimator considering simulations with sample size varying as $n = 200, 400, 600, 800$ and 1000. These quantities were calculated using the following formulas:

$$\text{Bias}(\theta) = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_r - \theta\right) \tag{3.3}$$

and

$$\text{MSE}(\theta) = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_r - \theta\right)^2, \tag{3.4}$$

where $R = 5000$ is the number of replications used for each sample size in the simulations and $\hat{\theta}_r$ is the estimate of θ in the r-th replicate of size n. For the calculations using the formulas above, the true values assumed for θ are necessary and they are given in Table 1.

The bias of the estimators was estimated using Equation (3.3) under each of the four scenarios of simulation described in Subsection 3.2.2. The results are summarized in Figure 1, which was made taking as reference the values printed in Table 10 of Appendix B. The graphs of Figure 1 allow to conclude that, for the four CFA-SMN models considered in the simulations, the estimators' bias appears to decrease with increasing sample sizes. This tendency is more markedly seen for the estimators of the parameters comprising the error's variance matrix and common factors' covariance matrix. The bias of estimators related to the parameters in the loading matrix appears to approach zero, but not so fast as happens with the remaining parameters.

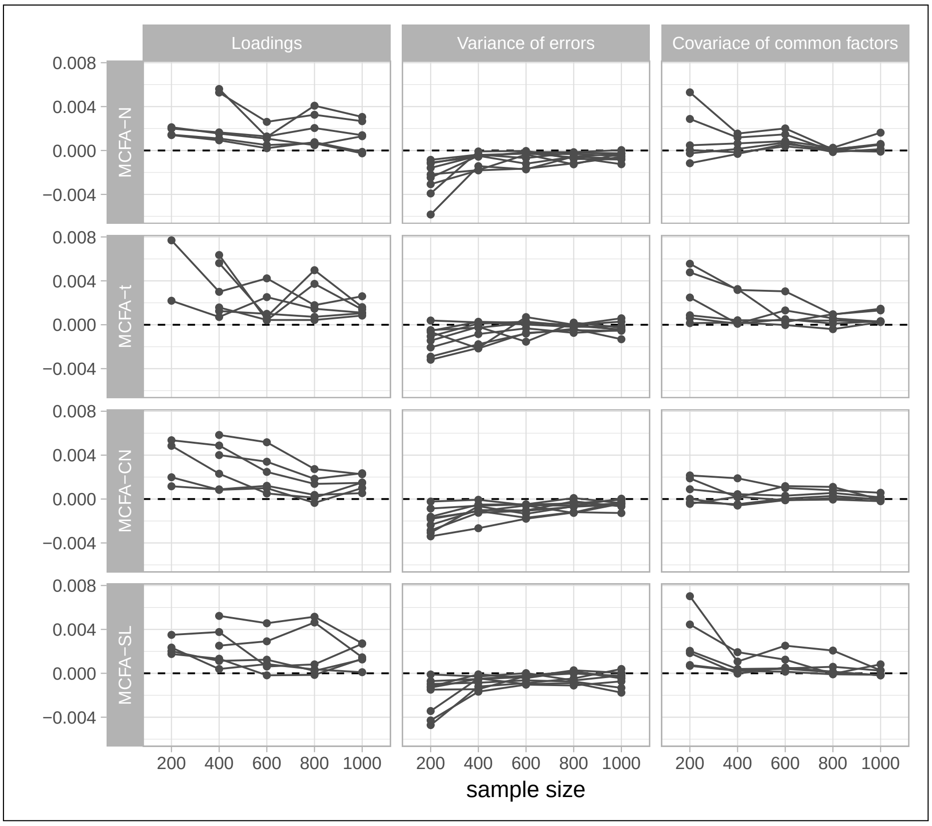


Figure 1: Bias

It is important to call attention to some points that were not plotted in the graphs of Figure 1, relating to the bias of the estimators of loading parameters. That happens only for results os simulations with sample size equals to 200. Concerning only this sample size, for the scenarios of simulation involving the MCFA-N, MCFA-CN and MCFA-SL models, the loading parameters λ_{83} and λ_{93} are not plotted in Figure 1. The same happens for the parameters λ_{52}, λ_{62}, λ_{83} and λ_{93} for the simulations involving the MCFA-t model. The bias associated with those parameters can be seen in Table 10.

Their estimation lead to large bias in samples of size 200. They were omitted in order to facilitate the interpretation of the graphs in Figure 1. However, that peculiar result points out possible issues for estimation of MCFA-SMN models in samples of size as small as 200 observations per group $g = 1, \ldots, G$. Although, further studies are needed to fully understand the effect of sample size in parameter estimation in the context of MCFA-SMN models.

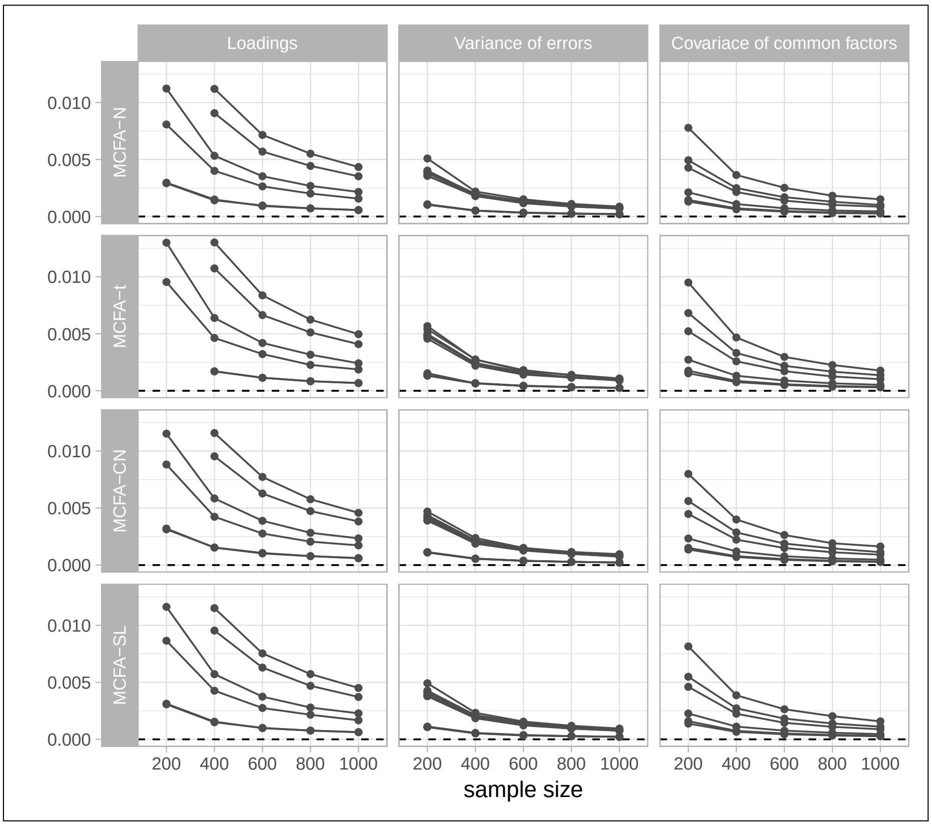


Figure 2: Mean square error (MSE)

The simulation results concerning the MSE of the estimators are summarized in Figure 2, which is based on the values printed in Table 11 of Appendix B. From the graphs of Figure 2 it can be asserted that the estimators' MSE decreases with increasing sample size. Again, it has a lower decrease for the loading parameters. As happened with the estimators' bias, in the simulations with samples of size equals to 200 some loading parameters presented high MSE. Those parameters are the same ones pointed out earlier in the text when analyzing the estimators' bias. Altogether, the results of Figure 2 give support to the consistency of estimators obtained in our study, a desired property since the proposed estimators are maximum likelihood estimators.

3.3.2 Standard errors and confidence intervals

For studying the behavior of standard errors of estimates obtained with the estimators of Section 2.4 we considered the CDM and EFI methods for approximation of the Fisher information. For each sample size $n = 200, 400, 600, 800$ and 1000 in the simulations, the average standard errors obtained through the CDM and EFI methods were calculated. The results were compared with the Monte Carlo standard error, which was calculated using the formula below:

$$\text{MC SE}(\hat{\theta}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_r^2 - \left(\frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_r\right)^2}, \tag{3.5}$$

where $R = 5000$ is the number of replications used for each sample size in the simulations and $\hat{\theta}_r$ is the estimate of θ in the r-th replicate of size n.

In Section 2.5 it was proposed two ways of obtaining standard errors for estimates of parameters in MCFA-SMN models. Both methods were based on approximations of the Fisher information. The EFI method proposes as standard errors the square root of the diagonal elements of the empirical Fisher information matrix, while the CDM method makes it using a numerical approximation of the Fisher information. For each scenario and sample size in our simulation study the standard errors obtained through the EFI and CDM methods were assessed in all 5000 replications and its average value were calculated and retained. Tables 13 and 14 of Appendix B bring those average values based on the EFI and CDM methods, respectively.

In order to evaluate the performance of the standard errors obtained through both methods we compared their average in the simulations with the Monte Carlo standard errors calculated using Equation (3.5). Figure 3 pictures the behavior of the Monte Carlo standard errors under estimation in the four MCFA-SMN models considered in our study. It can be seen that the standard errors of estimates of all parameters became smaller when sample size is increased. That consistency property is desired for the proposed estimator and reassure the results for the estimators' MSE described earlier in the text. As occurred for the bias and MSE, some parameters in all MCFA-SMN models study showed high Monte Carlo standard errors in simulations with samples of size 200. The points corresponding to high Monte Carlo standard errors are omitted in Figure 3, it corresponds exactly to the same loading parameters described earlier in the last section of the text.

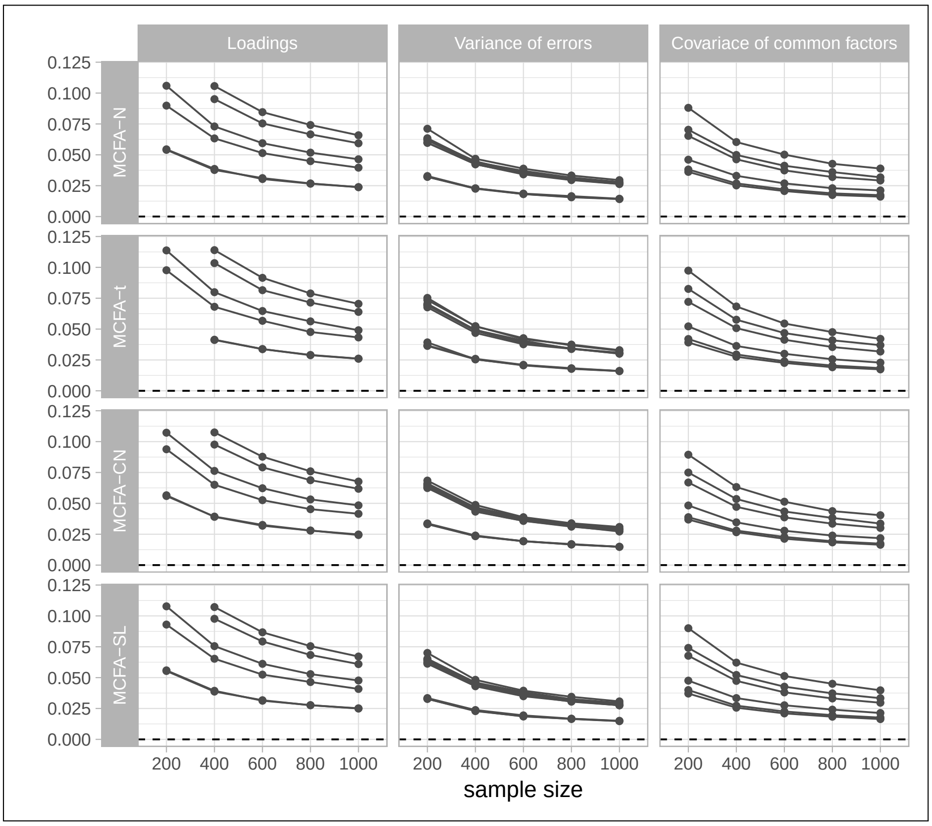
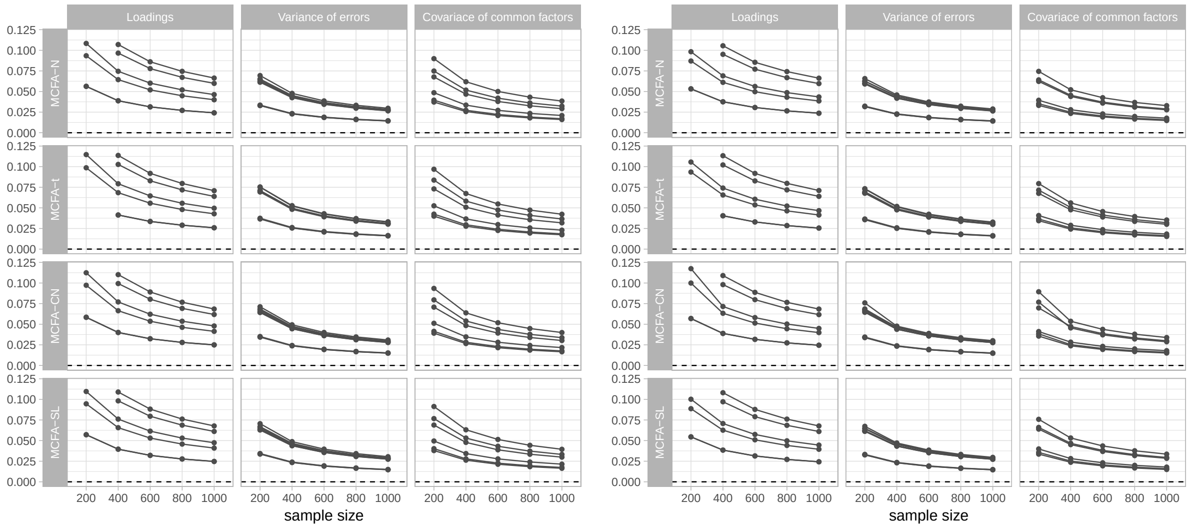


Figure 3: Monte Carlo standard error (MCSE)

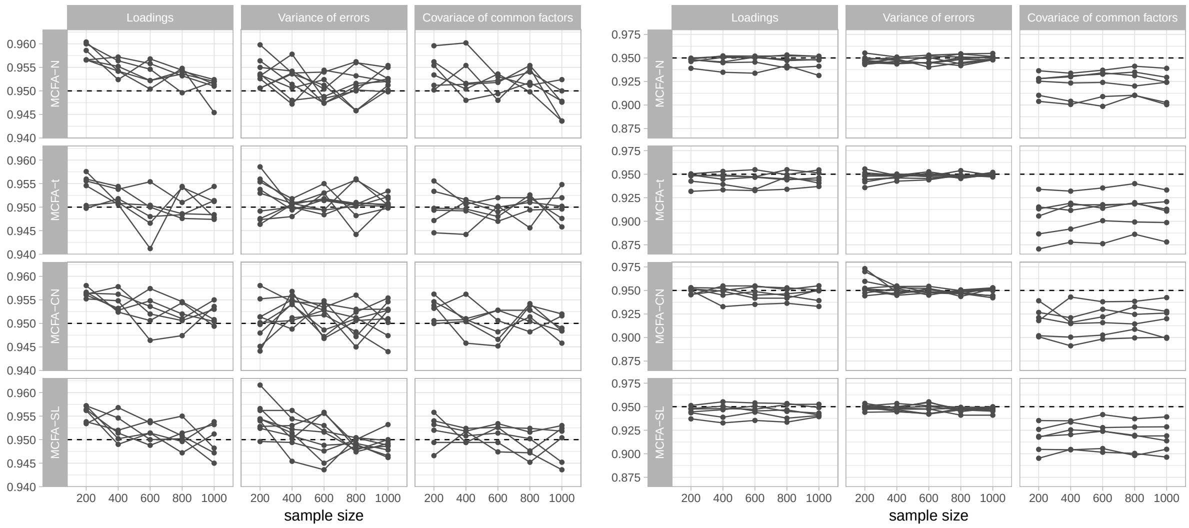
The Monte Carlo standard errors serve as reference to study the adequacy of the EFI and CDM as methods for generating standard errors for estimates of parameters in MCFA-SMN models. It is expected that those standard errors behave similarly to the ones generated with the Monte Carlo method, verifying its approximate magnitude and the consistency property. Figures 4a and 4b depict the average standard errors printed in Tables 13 and 14 of Appendix B, which refer to the EFI and CDM methods, respectively. It can be seen in the Tables 13 and 14 that the magnitudes of the average standard errors generated through the EFI method are very alike to the Monte Carlo standard errors, although the CDM methods lead to standard errors somewhat smaller then the Monte Carlo standard errors. The consistency property is observed for the standard errors obtained through either method, EFI or CDM.

In accordance with the high MSE estimated for the loading parameters λ_{83} and λ_{93}, the standard errors for this parameters were also higher relatively to the others loadings, considering the Monte Carlo, EFI and CDM standard errors. It can be observed in Tables 12, 13 and 14 of Appendix B that the set of loading parameters associated with the second common latent factor, namely, λ_{52} and λ_{62}, showed the smaller standard errors. Among the parameters comprising the covariance matrix of common latent factors, the variances ς_{11}, ς_{22} and ς_{33} showed higher standard errors, calculated using any of the three methods Monte Carlo, EFI or CDM.



(a) EFI (b) CDM

Figure 4: Empirical Fisher information (EFI) and central difference method (CDM) standard errors.



(a) EFI (b) CDM

Figure 5: Empirical Fisher information (EFI) and central difference method (CDM) 95% confidence intervals.

The probability coverage of 95% confidence intervals (CI) based on the standard errors generated through the EFI and CDM methods are shown in Figures 5a and 5b.

It can be seen that the EFI method leads to CI close to the nominal level of 95% confidence, while the lower standard errors of the CDM method lead to CI with confidence under the nominal level, mainly for the parameters in the covariance matrix of common factors. The lower coverage of CI constructed using the CDM method can be attributed to its respective smaller standard errors. To understand better this problem it is advised a revision of the computer codes implemented to calculate the standard errors through the CDM method.

4 Application

4.1 Resumo da seção

Neste capítulo apresentamos uma aplicação dos modelos MCFA-SMN. A aplicação é no campo da genética médica e utiliza dados reais de expressção gênica em pacientes com câncer de pâcreas. Nós propomos um modelo de análise fatorial exploratório para múltiplos grupos a partir da expressão de 11 proteínas envolvidas na regulação da matrix extracelular e do meio onde células tumorais se desenvolvem. Em seguida, propomos um modelo confirmatório basado nos resultados da modelagem anterior. Tanto na abordagem exploratória quanto na confirmatória, quatro modelos MCFA-SMN são estimados, supondo-se fatores latentes com distribuição normal, t-*Student*, normal contaminada e *slash*. O modelo confirmatório final é interpretado sob a luz de descobertas científicas recentes na área de oncologia a respeito das moléculas inseridas no contexto da modelagem estatística.

4.2 Context of application

In oncology, the study of biological pathways involved in the regulation of gene expression in tumor cells has been a central issue for understanding the pathological dynamic of cancer. A biological pathway is a cascade of chemical and physical events connecting molecules and cells in complex networks for the control of physiological functions. In cancer, those networks are altered due to genetic changes that affects the expression of genes in tumor cells (PONDER, 2001).

Since the expression of genes is intimately related to the concentration of its associated polypeptide in the circulatory system, the measurement of proteins in the blood and serum became the basic tool for gathering data towards the analysis and modeling of biological pathways (IACOB *et al.*, 2016). Recently, the availability of powerful technologies of high-throughput genomic data, *e.g.* RNA-seq and microarray, made it possible to simultaneously analyze all molecules composing the transcriptome of a cell, *i.e.* the set of all RNA molecules in the cell (PHAM *et al.*, 2016).

As a result, scientists in the field of Medicine are now turning their attention to the analysis of gene co-expression (IACOB *et al.*, 2016; PHAM *et al.*, 2016). Gene co-expression refers to the interrelationship of genes for the co-regulation of their expression levels, *i.e.* genes interacting together to activate or deactivate their transcription into RNA in the cell (ROY *et al.*, 2014).

4.3 Data set

In the following we shall present and analyze a data set from the field of oncology. The variables comprising the data set are gene expressions measured with microarrays. The data set includes observations of two independent populations of tumor cells coming from patients with pancreas cancer, where it is known the true allocation of observations in each population. The two samples are denoted, respectively, as TCGA (with sample size $n = 146$) and ICGCMICRO (with sample size $n = 265$). The TCGA and ICGCMICRO data sets are of public domain and can be freely accessed from the *bioconductor* platform through the software R, by calling the package *MetaGxPancreas*.

The data set selected has measurements on the expression of hundreds of distinct genes, from which we have selected 11 target genes for our analysis based on the current scientific knowledge on the dynamic of cancer cells in its microenvironment, the extracellular matrix (CASEY *et al.*, 2007; LI *et al.*, 2013; FANG *et al.*, 2014; GIALELI *et al.*, 2014; COX *et al.*, 2015; GASCARD and TLSTY, 2016; JIA *et al.*, 2016; HAMMER *et al.*, 2017). The proteins regulated by the chosen genes are listed in Table 2 below.

Table 2: Name of proteins associated with the 11 targeted genes.

Abbreviation	**Protein**
COL3A1	Collagen, type III, alpha 1
COL10A1	Collagen, type X, alpha 1
COL11A1	Collagen, type XI, alpha 1
COL5A2	Collagen, type V, alpha 2
THBS2	Thrombospondin 2
PLAU	Plasminogen activator, urokinase
PDGFRA	Platelet-derived growth factor receptor, alpha polypeptide
PDGFRB	Platelet-derived growth factor receptor, beta polypeptide
ACTA2	Actin, alpha 2
TIMP3	IMP metallopeptidase inhibitor 3
IGF1	Insulin-like growth factor 1 (somatomedin C)

Usually, data gathered in studies involving the measurement of gene expression is available in the $\log_2$-transformed scale, this is true for several platforms commonly used for processing biological molecules and assessment of gene expression levels (TENG *et al.*, 2013) (including the Affymetricx plataform, which was used to obtain the data set under analysis in this section). Our first analytic decision was to investigate the data set in its original scale, hence we transformed back the data applying the inverse of the logarithmic function with base 2. The next step was to standardize the data by subtracting from each observed variable its sample mean and dividing it by its sample standard deviation, a commonly used strategy for data analysis using factor analytic models (MEHRA, 1973). Figures 6 and 7 show the sample distribution of each raw gene expression after standardization.

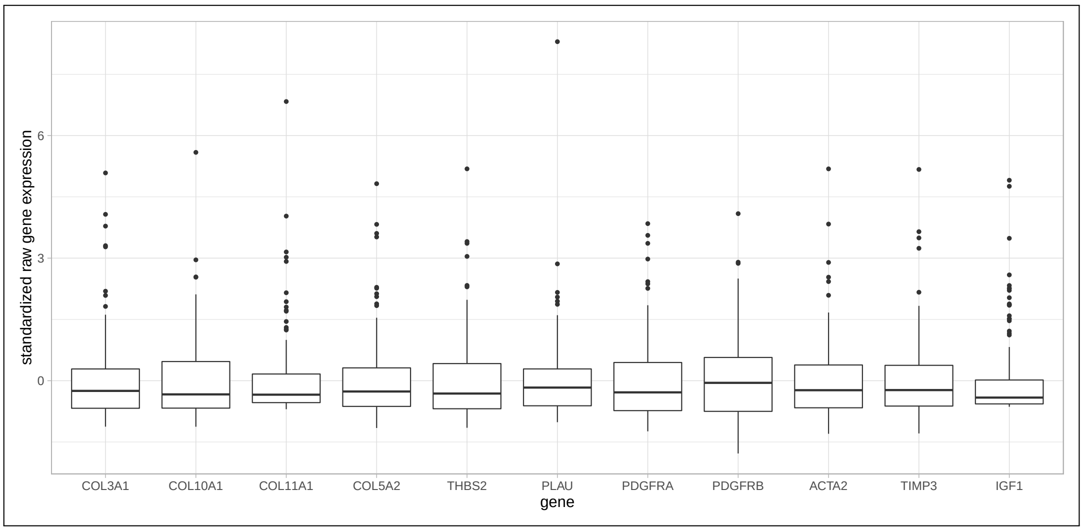


Figure 6: Standardized raw variables for the TCGA data set.

The two groups, TCGA and ICGCMICRO, have similar marginal distributions for the standardized raw gene expressions. The presence of outliers in almost all variables occurs in both groups, suggesting the need for probabilistic models with heavier tails than the normal distribution. However, it can be seen a clear sign of right skewness in both data sets.

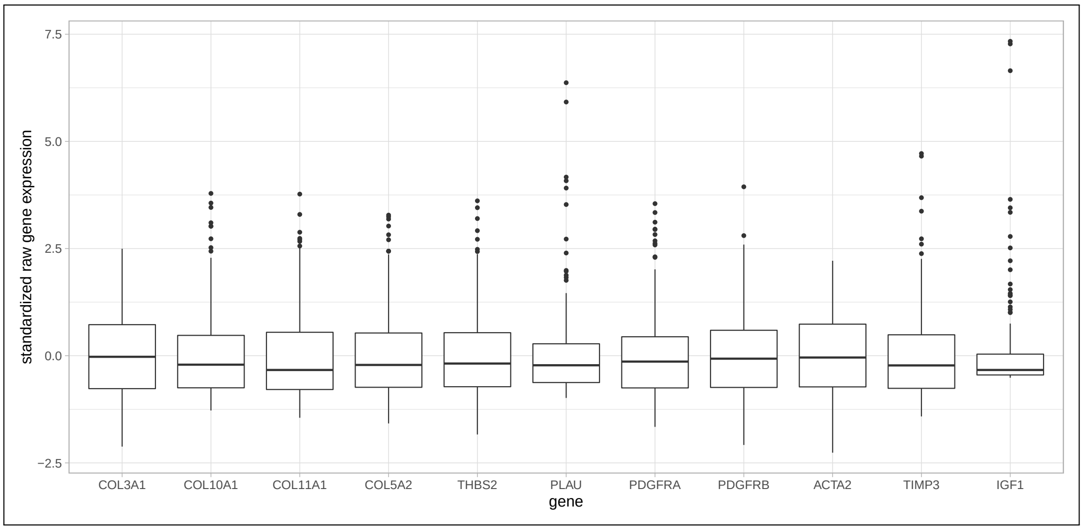


Figure 7: Standardized raw variables for the ICGCMICRO data set.

An important issue in factor analysis is the determination of the number of common factors. A commonly used criterion for that aim is the rule of Kaiser that says to retain only common factors whose eigenvalues are greater than one (KAUFMAN and DUNLAP, 2000). Figure 8 shows the eigenvalues of the sample covariance matrices observed for the TCGA and ICGCMICRO data sets. According to the Kaiser rule it should be retained only two common factors in each of the groups, TCGA and ICGCMICRO.

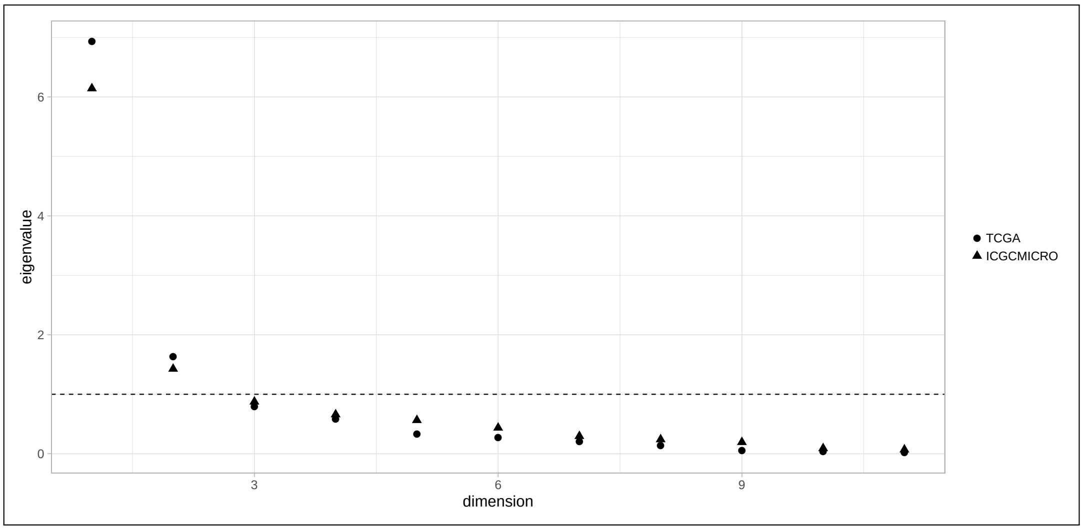


Figure 8: Eigenvalues of the sample covariance matrices for the TCGA and ICGCMICRO data sets.

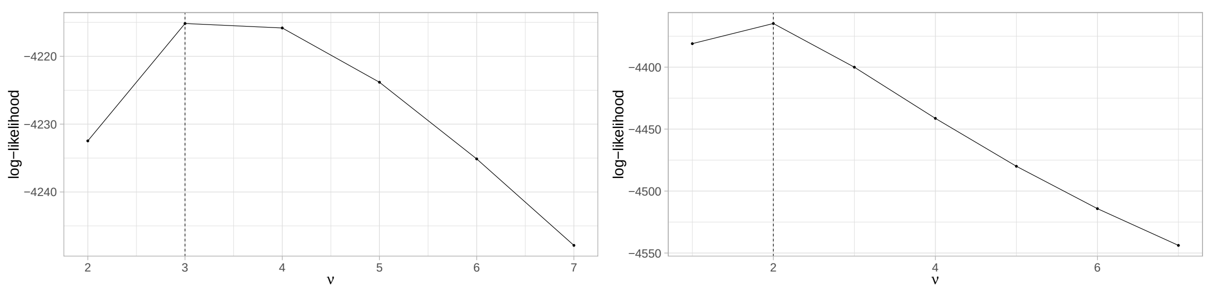
4.4 Exploratory model

Our fist attempt in modeling the covariance structure of the observed data set was in estimating a exploratory multiple group factor analysis model, assuming the observed variables follow a probability distribution in the SMN class. That action allowed us to investigate the general behavior of the loading parameters without the need to assume any prior knowledge about its configuration, except for the number of common latent factors, which was fixed and equals to two (according to the criterion of Kaiser). The model matrices are defined in 4.1, where each parameter is explicitly shown.

The identification of parameters in the exploratory model can be easily verified with the help of Theorem 1 and Proposition 1. According to this theoretical results a MCFA-SMN model having loading matrices with a *triangular matrix of zeros* form and covariance matrices of common factors equal to the identity matrix will be identified as long as the parameters $\lambda_{j,j}^{(g)}$, $j = 1,2$ and $g = 1,2$, are identified. In this application we fixed the parameters $\lambda_{1,1}^{(g)}$ and $\lambda_{2,2}^{(g)}$ equals to 1, $g = 1,2$.

$$\mathbf{\Lambda}_g = \begin{bmatrix} 1 & 0 \\ \lambda_{2,1} & 1 \\ \lambda_{3,1} & \lambda_{3,2} \\ \lambda_{4,1} & \lambda_{4,2} \\ \lambda_{5,1} & \lambda_{5,2} \\ \lambda_{6,1} & \lambda_{6,2} \\ \lambda_{7,1} & \lambda_{7,2} \\ \lambda_{8,1} & \lambda_{8,2} \\ \lambda_{9,1} & \lambda_{9,2} \\ \lambda_{10,1} & \lambda_{10,2} \\ \lambda_{11,1} & \lambda_{11,2} \end{bmatrix}, \ \boldsymbol{\zeta}_g = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \mathbf{\Psi}_g = \text{diag}(\psi_{j,j})_{j=1}^{11}, \ g = 1,2. \tag{4.1}$$

In order to choose for the observed variables a particular probability distribution in the SMN class we estimated MCFA-N, MCFA-t, MCFA-CN and MCFA-SL models. For the MCFA-t and MCFA-SL the parameter ν, that is assumed fixed in our model definition, was chosen by evaluation of the profile log-likelihood of ν in both models. Figure 9 shows the results obtained. Hence, for the MCFA-t model the value of $\nu = 3$ leads to the highest log-likelihood. In the case of MCFA-SL model the value of $\nu = 2$ is the one leading to the highest log-likelihood.



(a) MCFA-t (b) MCFA-SL

Figure 9: Profile log-likelihood of ν in MCFA-t and MCFA-SL exploratory models.

For the MCFA-CN model, the parameters ξ and γ where also chosen by profiling the log-likelihood at a grid of values ranging from 0.10 to 0.40 at spaces of size 0.05. This choice of grid was made assuming parameter values with two significant decimal places would be close enough estimates to the actual parameter value in the population. Although a larger grid was tried in the application, we show only the range of values previously mentioned. The results are shown Table 3. Hence, $\xi = 0.20$ and $\gamma = 0.20$ leads to the largest log-likelihood in the chosen grid.

Table 3: Log-likelihood for a grid of values of ξ and γ in the MCFA-CN exploratory model

ξ \ γ	0.10	0.15	0.20	0.25	0.30	0.35	0.40
0.10	-4381.948	-4379.322	-4382.915	-4390.685	-4401.743	-4415.642	-4432.173
0.15	-4352.235	-4344.662	-4344.160	-4348.124	-4355.450	-4365.644	-4378.506
0.20	-4351.172	-4341.133	***-4338.457***	-4340.448	-4345.945	-4354.378	-4365.468
0.25	-4366.603	-4355.433	-4351.738	-4352.777	-4357.343	-4364.833	-4374.946
0.30	-4391.569	-4380.026	-4375.901	-4376.470	-4380.524	-4387.455	-4396.961
0.35	-4422.000	-4410.514	-4406.269	-4406.609	-4410.349	-4416.891	-4425.938
0.40	-4455.405	-4444.238	-4440.043	-4440.271	-4443.779	-4449.986	-4458.602

Including the MCFA-N model, we have chosen, based on the results of Figure 9 and Table 3, to maintain the models MCFA-t($\nu = 3$), MCFA-SL($\nu = 2$) and MCFA-CN($\xi = 0.20$, $\gamma = 0.20$). Table 4 shows the point estimates and estimates of standard errors for the parameters in the four models. The standard errors were estimated using the method of Meilijson (1989) based on the empirical Fisher information matrix.

Table 4: Point estimates and standard errors (in parenthesis) for the parameters in the exploratory models: MCFA-N, MCFA-t($\nu = 3$), MCFA-CN($\xi = 0.20, \gamma = 0.20$) and MCFA-SL($\nu = 2$).

Parameters	MCFA-N		MCFA-t		MCFA-CN		MCFA-SL	
$\lambda_{2,1}$	0.831	(0.144)	0.792	(0.178)	0.763	(0.193)	0.766	(0.242)
$\lambda_{3,1}$	0.959	(0.038)	0.825	(0.041)	0.842	(0.038)	0.822	(0.042)
$\lambda_{4,1}$	1.060	(0.027)	0.981	(0.026)	0.958	(0.027)	0.947	(0.026)
$\lambda_{5,1}$	0.990	(0.031)	0.936	(0.032)	0.903	(0.037)	0.895	(0.035)
$\lambda_{6,1}$	0.490	(0.077)	0.529	(0.051)	0.474	(0.060)	0.481	(0.055)
$\lambda_{7,1}$	0.456	(0.133)	0.543	(0.164)	0.421	(0.172)	0.436	(0.208)
$\lambda_{8,1}$	0.706	(0.089)	0.765	(0.097)	0.673	(0.098)	0.678	(0.113)
$\lambda_{9,1}$	0.716	(0.063)	0.876	(0.055)	0.796	(0.055)	0.792	(0.057)
$\lambda_{10,1}$	0.808	(0.052)	0.772	(0.061)	0.709	(0.064)	0.715	(0.070)
$\lambda_{11,1}$	0.161	(0.110)	0.239	(0.094)	0.166	(0.100)	0.161	(0.114)
$\lambda_{3,2}$	-0.150	(0.068)	-0.161	(0.050)	-0.105	(0.047)	-0.122	(0.047)
$\lambda_{4,2}$	0.084	(0.052)	0.027	(0.039)	0.051	(0.037)	0.031	(0.035)
$\lambda_{5,2}$	0.150	(0.048)	0.102	(0.041)	0.138	(0.037)	0.101	(0.036)
$\lambda_{6,2}$	-0.348	(0.135)	-0.180	(0.070)	-0.205	(0.077)	-0.176	(0.063)
$\lambda_{7,2}$	1.140	(0.088)	1.160	(0.077)	1.080	(0.068)	1.040	(0.068)
$\lambda_{8,2}$	0.595	(0.075)	0.583	(0.076)	0.531	(0.068)	0.488	(0.067)
$\lambda_{9,2}$	0.242	(0.087)	0.195	(0.069)	0.179	(0.066)	0.164	(0.062)
$\lambda_{10,2}$	0.359	(0.064)	0.368	(0.055)	0.343	(0.051)	0.315	(0.051)
$\lambda_{11,2}$	0.717	(0.121)	0.505	(0.062)	0.486	(0.065)	0.459	(0.057)
$\psi_{1,1}$	0.193	(0.013)	0.117	(0.010)	0.107	(0.009)	0.085	(0.007)
$\psi_{2,2}$	0.559	(0.060)	0.418	(0.049)	0.369	(0.042)	0.330	(0.041)
$\psi_{3,3}$	0.197	(0.018)	0.114	(0.011)	0.106	(0.009)	0.082	(0.007)
$\psi_{4,4}$	0.037	(0.007)	0.025	(0.005)	0.019	(0.004)	0.016	(0.003)
$\psi_{5,5}$	0.144	(0.008)	0.100	(0.008)	0.091	(0.007)	0.070	(0.005)
$\psi_{6,6}$	0.735	(0.031)	0.296	(0.020)	0.305	(0.013)	0.193	(0.011)
$\psi_{7,7}$	0.222	(0.054)	0.121	(0.040)	0.111	(0.033)	0.068	(0.027)
$\psi_{8,8}$	0.407	(0.028)	0.330	(0.032)	0.263	(0.021)	0.214	(0.017)
$\psi_{9,9}$	0.529	(0.019)	0.317	(0.025)	0.283	(0.017)	0.210	(0.014)
$\psi_{10,10}$	0.376	(0.024)	0.236	(0.019)	0.206	(0.014)	0.160	(0.011)
$\psi_{11,11}$	0.739	(0.040)	0.264	(0.019)	0.283	(0.014)	0.178	(0.011)

Most of the parameters comprising the loading matrix have similar point estimates under all four models. The major differences in point estimates among MCFA-SMN models are in the parameters of the variance matrix of errors (or, equivalently, specific latent factors). For this parameters the MCFA-SL model showed the smallest point estimates. The standard errors are smaller for most parameters in the models MCFA-t, MCFA-CN and MCFA-SL, again specially for the variance of errors. In the context of application of CFA models using robust estimators, Zhong and Yuan (2011) also observed this behavior of point estimates and standard errors' estimate when comparing the normal CFA model against robust CFA models.

We compared the fit of the four estimated models using the Akaike Information Criterion (AIC). The AIC is frequently used for model selection in FA (AKAIKE, 1987; CASTRO *et al.*, 2014). Table 5 shows the AIC for each of the four models, the smallest AIC being associated to the MCFA-t($\nu = 3$). Then, using the AIC as a measure of model selection, we can consider the MCFA-t($\nu = 3$) model as presenting the best fit among the estimated models for the data set analyzed.

Table 5: AIC values for the four fitted exploratory models: MCFA-N, MCFA-t($\nu = 3$), MCFA-CN($\xi = 0.20, \gamma = 0.20$) and MCFA-SL($\nu = 2$).

MCFA-N	MCFA-t	MCFA-CN	MCFA-SL
9558.643	***8490.332***	8736.913	8789.535

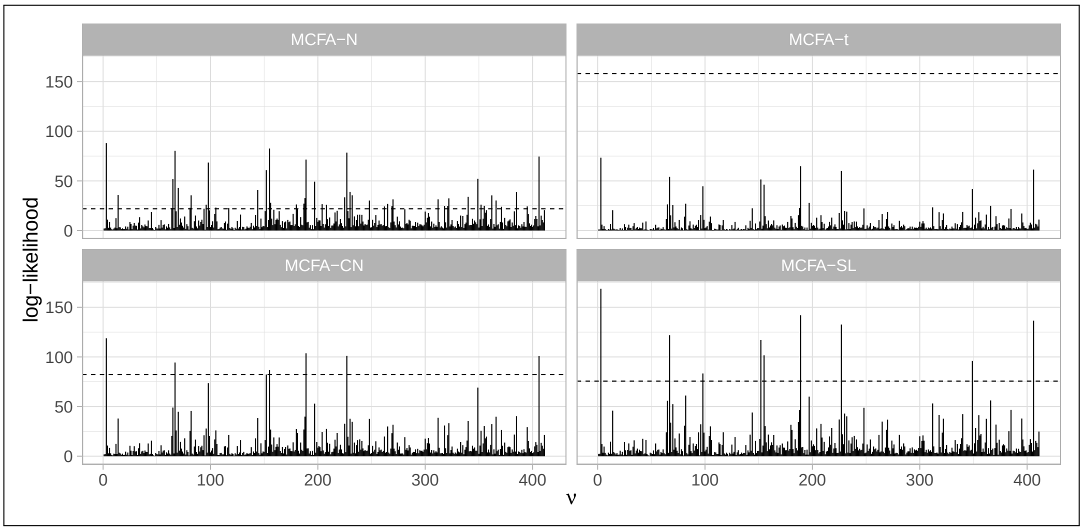


Figure 10: Mahalanobis distances for each of the four estimated exploratory models. Dotted line indicates the 97.5% quantile of the appropriate Mahalanobis distances distribution according to the response variable distribution.

An important aspect of the MCFA-SMN models fitted is the behavior of its associated Mahalanobis distances. Figure 10 shows the Mahalanobis distances for the four estimated models, together with its 97.5% theoretical quantile calculated using the results of Subsection 2.4.1. It can be seen that for the MFCA-SMN model there are more distances greater than the theoretical quantile, while for the remaining SMN distributions the Mahalanobis distances stay mainly confined to the limits established by the probability theory. Although, only the MCFA-t ($\nu = 3$) model shows Mahalanobis distances systematically smaller than the distances observed for the MCFA-N model.

Taking the MCFA-t ($\nu = 3$) model as the most appropriate to describe the covariance structure of the 11 gene expressions targeted in this analysis, we finally consider an orthogonal rotation of the loading matrices in order to interpret the results obtained. Table 6 shows the rotated loading matrix common to both groups, TCGA and ICGCMICRO, with the rotation being made according to the Varimax criterion. The interpretation of results is differed to Subsection 4.6.

Table 6: Varimax rotation of the estimated loading matrix for the MCFA-t($\nu = 3$) exploratory model.

Genes	1-th Common Factor	2-th Common Factor
COL3A1	**0.947**	0.321
COL10A1	0.430	**1.200**
COL11A1	**0.833**	0.112
COL5A2	**0.920**	0.340
THBS2	**0.854**	0.397
PLAU	**0.559**	-0.001
PDGFRA	0.142	**1.270**
PDGFRB	0.538	**0.798**
ACTA2	**0.767**	0.466
TIMP3	**0.614**	0.596
IGF1	0.065	**0.555**

4.5 Confirmatory model

The exploratory multiple factor analysis model fitted in Subsection 4.4 reveled an underlining latent structure capable of explaining the covariaces between observed indicators (or, equivalently, observed variables) in terms of only two common factors. Those common factors are related to the observed indicators through a matrix of loadings, which, after rotation (Table 6), revels a clear pattern of association between indicators: with the proteins COL3A1, COL11A1, COL5A2, THBS2, PLAU, ACTA2 and TIMP3 showing bigger weights in the first common factor and the proteins COL10A1, PDGFRA, PDGFRB and IGF1 with bigger weights in the second common factor.

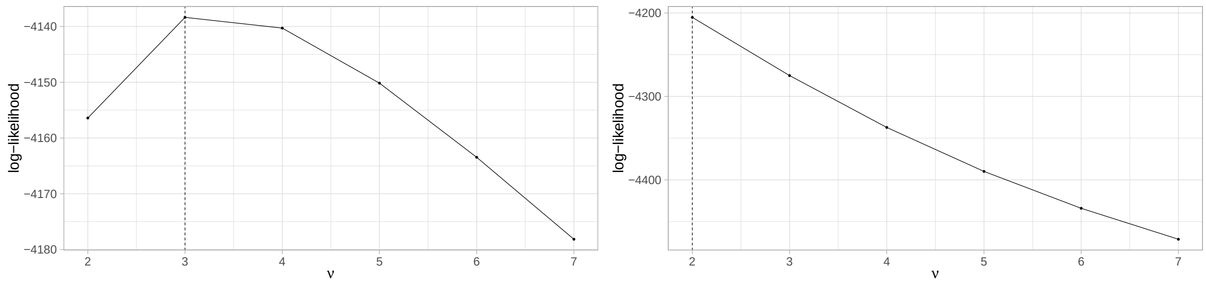
Although, by fixing the covariance matrix of common latent factors equal to the identity matrix, the exploratory model assumes orthogonal common latent factors, leading to less flexible data analysis. Hence, a more flexible model allowing for nonzero covariance between common factors would permit to investigate the existence of oblique common factors. Below we propose a confirmatory multiple factor analysis model where the loading matrices follow a simple structure (as discussed in Subsection 2.3) and the covariance matrices of common factors have a free parameter for estimation. In this confirmatory model, the scale of the indicators is set up by fixing the variance of each common factor equals to one.

$$\mathbf{\Lambda}_g = \begin{bmatrix} \lambda_{1,1} & 0 \\ 0 & \lambda_{2,2} \\ \lambda_{3,1} & 0 \\ \lambda_{4,1} & 0 \\ \lambda_{5,1} & 0 \\ \lambda_{6,1} & 0 \\ 0 & \lambda_{7,2} \\ 0 & \lambda_{8,2} \\ \lambda_{9,1} & 0 \\ \lambda_{10,1} & 0 \\ 0 & \lambda_{11,2} \end{bmatrix}, \ \zeta_g = \begin{bmatrix} 1 & \varsigma_{2,1} \\ \varsigma_{2,1} & 1 \end{bmatrix}, \ \mathbf{\Psi}_g = \text{diag}(\psi_{j,j})_{j=1}^{11}, \ g = 1,2. \tag{4.2}$$

For this application the parameter identification can be verified by searching the conditions of Theorem 2. According to this theorem the Jacobian matrix R_2 defined in Equation 2.10 must be of full column rank. For the proposed model, the associated matrix R_2 has 12 columns and is of full column rank, since a simple calculation leads to rank$(R_2) = 12$. Hence, the MCFA-SMN models to be proposed next will have all its parameters identified.

As in the previous section, the interest resides in the estimation of four MCFA-SMN models: MCFA-N, MCFA-t, MCFA-CN and MCFA-SN. In order to determine the fixed values for the parameter ν appearing in the MCFA-t and MCFA-SL models and the fixed value of the parameters ξ and γ associated to the MCFA-CN model, we obtained the profile log-likelihood of this parameters. The profiles are shown in Figure 11 for the

MCFA-t and MCFA-SL models and in Table 7 for the MCFA-CN model.



(a) MCFA-t (b) MCFA-SL

Figure 11: Profile log-likelihood of v in MCFA-t and MCFA-SL confirmatory models.

According to the profiles, the MCFA-t($v = 3$), MCFA-SL($v = 2$) and MCFA-CN($\xi = 20$, $\gamma = 0.30$) lead to higher log-likelihoods. Comparing with the fitted exploratory models, here only the parameter γ in the MCFA-CN model resulted differently. The parameter estimates of the selected models are shown in Table 8.

Table 7: Log-likelihood for a grid of values of ξ and γ in the MCFA-CN confirmatory model

ξ \ γ	0.10	0.15	0.20	0.25	0.30	0.35	0.40
0.10	-4299.290	-4290.560	-4289.920	-4294.350	-4302.480	-4313.510	-4326.790
0.15	-4262.380	-4249.940	-4245.560	-4245.660	-4248.250	-4251.520	-4254.840
0.20	-4261.450	-4246.600	-4239.570	-4236.760	***-4236.390***	-4237.610	-4240.340
0.25	-4279.430	-4263.390	-4255.170	-4251.340	-4250.390	-4251.610	-4254.770
0.30	-4308.020	-4291.640	-4283.090	-4279.080	-4278.190	-4279.710	-4283.360
0.35	-4342.610	-4326.370	-4317.890	-4314.000	-4313.300	-4315.120	-4319.120
0.40	-4380.400	-4364.600	-4356.370	-4352.670	-4352.180	-4354.200	-4358.390

In Table 8 it can be seen that the four MCFA-SMN models reveal a strong correlation between the two common factors (with estimates above 0.90 in all models). Also the correlation parameter $\varsigma_{2,1}$ had similar estimates for all models. Otherwise, the remaining parameters showed greater differences in estimates between models, specially when comparing the MCFA-N model with the other three models. As occurred with the exploratory models, the models assuming heavier tails for the distribution of observed variables showed smaller standard errors mainly for the estimated variances of errors.

Table 8: Point estimates and standard errors (in parenthesis) for the parameters in the confirmatory models: MCFA-N, MCFA-t($\nu = 3$), MCFA-CN($\xi = 0.20, \gamma = 0.30$) and MCFA-SL($\nu = 2$).

Parameters	MCFA-N		MCFA-t		MCFA-CN		MCFA-SL	
$\zeta_{2,1}$	0.928	(0.014)	0.923	(0.014)	0.904	(0.016)	0.907	(0.016)
$\lambda_{1,1}$	0.904	(0.038)	0.781	(0.050)	0.736	(0.034)	0.614	(0.031)
$\lambda_{3,1}$	0.874	(0.038)	0.645	(0.048)	0.661	(0.035)	0.535	(0.034)
$\lambda_{4,1}$	0.976	(0.034)	0.779	(0.051)	0.758	(0.034)	0.628	(0.032)
$\lambda_{5,1}$	0.923	(0.035)	0.751	(0.050)	0.725	(0.034)	0.604	(0.032)
$\lambda_{6,1}$	0.427	(0.061)	0.418	(0.042)	0.362	(0.044)	0.317	(0.032)
$\lambda_{9,1}$	0.679	(0.058)	0.717	(0.056)	0.625	(0.047)	0.550	(0.040)
$\lambda_{10,1}$	0.768	(0.040)	0.637	(0.050)	0.586	(0.038)	0.494	(0.035)
$\lambda_{2,2}$	0.833	(0.037)	0.682	(0.055)	0.660	(0.041)	0.553	(0.039)
$\lambda_{7,2}$	0.576	(0.064)	0.580	(0.056)	0.499	(0.053)	0.427	(0.043)
$\lambda_{8,2}$	0.739	(0.054)	0.692	(0.055)	0.624	(0.046)	0.530	(0.042)
$\lambda_{11,2}$	0.217	(0.070)	0.248	(0.055)	0.193	(0.058)	0.163	(0.048)
$\psi_{1,1}$	0.177	(0.011)	0.082	(0.008)	0.101	(0.008)	0.067	(0.005)
$\psi_{2,2}$	0.302	(0.029)	0.159	(0.019)	0.183	(0.020)	0.123	(0.014)
$\psi_{3,3}$	0.231	(0.016)	0.103	(0.010)	0.126	(0.010)	0.081	(0.007)
$\psi_{4,4}$	0.043	(0.006)	0.022	(0.004)	0.024	(0.004)	0.016	(0.003)
$\psi_{5,5}$	0.143	(0.009)	0.076	(0.007)	0.091	(0.007)	0.061	(0.005)
$\psi_{6,6}$	0.813	(0.014)	0.244	(0.018)	0.378	(0.010)	0.185	(0.010)
$\psi_{7,7}$	0.664	(0.040)	0.340	(0.032)	0.406	(0.029)	0.269	(0.020)
$\psi_{8,8}$	0.449	(0.031)	0.261	(0.029)	0.271	(0.025)	0.193	(0.018)
$\psi_{9,9}$	0.534	(0.018)	0.244	(0.022)	0.303	(0.016)	0.186	(0.012)
$\psi_{10,10}$	0.406	(0.026)	0.205	(0.019)	0.241	(0.017)	0.158	(0.011)
$\psi_{11,11}$	0.948	(0.027)	0.264	(0.020)	0.412	(0.014)	0.207	(0.011)

The AIC for the MCFA-N, MCFA-t($\nu = 3$), MCFA-SL($\nu = 2$) and MCFA-CN($\xi = 20$, $\gamma = 0.30$) are shown in Table 9. Here, as in the exploratory approach, the fitted model with lower AIC was the MCFA-t($\nu = 3$). Comparing the AIC for the fitted exploratory and confirmatory models it can be seen smaller AIC for the letter. Hence, based on the AIC criterion, the confirmatory MCFA-t($\nu = 3$) model would be selected among all fitted exploratory or confirmatory models.

Table 9: AIC values for the four fitted confirmatory models: MCFA-N, MCFA-t($\nu = 3$), MCFA-CN($\xi = 0.20, \gamma = 0.30$) and MCFA-SL($\nu = 2$).

MCFA-N	MCFA-t	MCFA-CN	MCFA-SL
9461.330	***8322.731***	8518.777	8456.607

The Mahalanobis distances based on the four fitted confirmatory models are shown in Figure12. Under the MCFA-t($\nu = 3$) the observations have Mahalanobis distance below the 97.5% quantile. For the MCFA-CN($\xi = 0.20, \gamma = 0.30$) and MCFA-SL($\nu = 2$) few observations have Mahalanobis distance crossing its correspondent 97.5% quantile, specially when compared to the MCFA-N model. Although, differently from the exploratory approach, the confirmatory MCFA-t($\nu = 3$) model showed several Mahalanobis distances a little bit higher then the Mahalanobis distances calculated for the MCFA-N model.

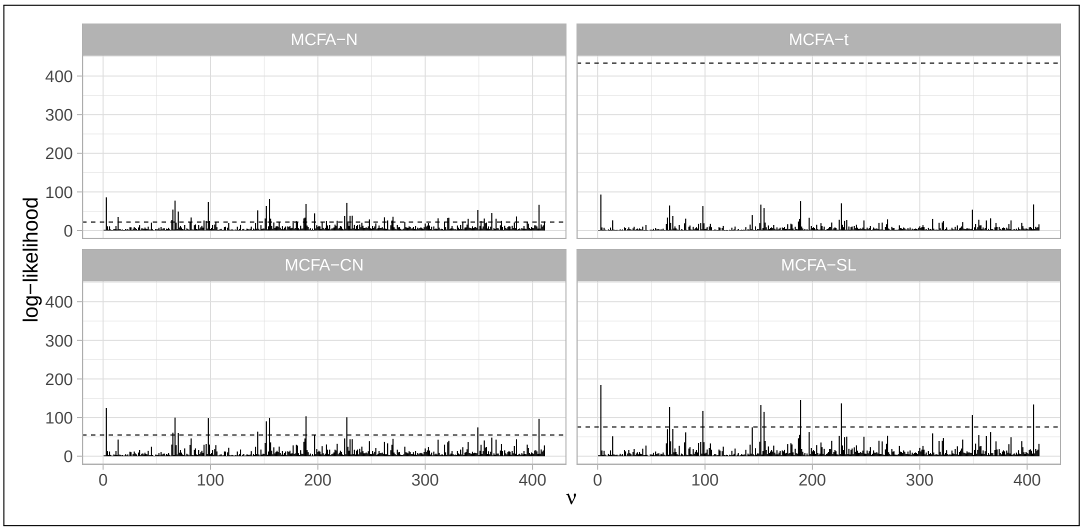


Figure 12: Mahalanobis distances for each of the four estimated confirmatory models. Dotted line indicates the 97.5% quantile of the appropriate Mahalanobis distances distribution according to the response variable distribution.

4.6 Interpretation of results

In Subsection 4.4 we defined and estimated MCFA-SMN models using an exploratory approach to determine a hypothesis relating the pattern of correlations between the 11 proteins of Table 2. This hypothesis led to confirmatory models (presented in Subsection 4.5) using loading matrices following a simple structure where the allocation of indicators in each column of the loading matrix was determined based on exploratory results. Now we shall argument the validity of our hypothesis by highlighting recent discoveries of laboratory researches on the molecular biology of tumor cells and the medium where this cells develop, the stroma and extra-cellular matrix (CASEY *et al.*, 2007; LI *et al.*, 2013; FANG *et al.*, 2014; GIALELI *et al.*, 2014; COX *et al.*, 2015; GASCARD and TLSTY, 2016; JIA *et al.*, 2016; HAMMER *et al.*, 2017).

The modern study of cancer has led scientists to reconsider the role of several biological factors involved in the dynamic of cancer with emphasis in the medium where tumor cells develop (GIALELI *et al.*, 2014). The tumor cell micro-environment is mainly composed of stroma and extra-cellular matrix, both composed by tissue, special types of cells and molecules located in the tumor cells' surrounds. Collagen is a central molecule in this micro-environment and has been shown to play a decisive part in tumor progression (GASCARD and TLSTY, 2016). Fang *et al.* (2014) discussed a series of biological pathways triggered by changes in tumor micro-environment and leading to tumor infiltration, angiogenesis, invasion and migration.

The linage of collagen molecules COL3A1, COL10A1, COL11A1, COL5A2 appears in gene signatures of several cancer types (MATONDO *et al.*, 2017). These molecules are listed in Table 2 and their gene expression represents the basic variables of our model. The remaining 7 proteins listed in Table 2 appear in our study to represent well known physiological relations among molecules present in the extracellular matrix.

Figure 13 gives a diagram explaining the basic relations between the 11 proteins chosen for composing the observable variables entering in our model. The diagram was made using the online software STRING, a popular software among medical scientists and clinicians used for gene annotation. The STRING software uses data from several databases for mining meaningful relationships among biological molecules. Although, the software is not directed to the study of data sets relating exclusively to cancer.

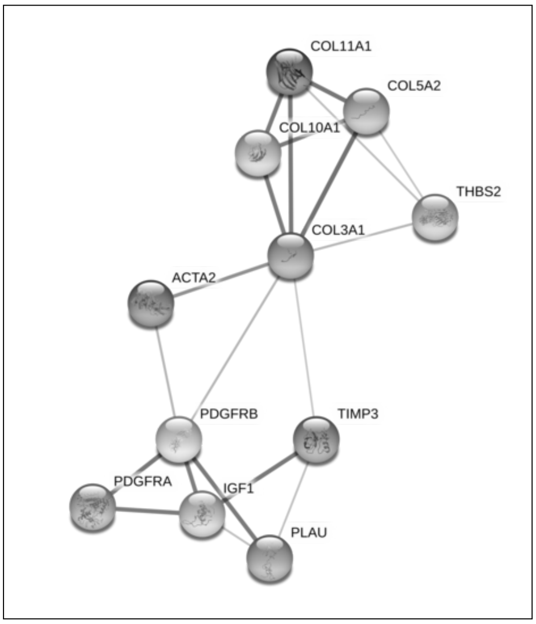


Figure 13: Diagram showing meaningful relationships for the 11 proteins regulated by targeted genes. The diagram was generated through the online software STRING directed to molecular biology and gene annotation. The width of the edges is directly proportional to strength of evidence of association between molecules

To improve our understanding on the relationship between the 11 targeted proteins in the context of cancer we have selected recent papers on the subject. An out standing discovery of Hammer *et al.* (2017) describes the connection between PDGFRA and collagen. According to laboratory experiments of Hammer *et al.* (2017), hyperactivation of PDGFRA leads to collagen deposition and decreasing in hydraulic permeability of collagen substrate, then contributing decisively in the dynamic of the extra-cellular matrix. Yet, this relationship connecting PDGFRA and collagen molecules is not present in the diagram of Figure 13. PDGFRA and PDGFRB are growth factors and represent important bio-markers of cancer (GIALELI *et al.*, 2014). The pathological functions of PDGFRB includes angiogenesis, metastase and proliferation of tumor cells (MATONDO *et al.*, 2017).

Another interesting molecule entering in our model is the protein PLAU. This protein converts plasminogen into plasmin, a substance able to degrade components of the extra-cellular matrix (LI *et al.*, 2013). Hence, PLAU activation facilitate the invasion of the extra-cellular matrix and stimulates angiogenesis (LI *et al.*, 2013), *i.e.* vascular development. Finally, PLAU mediates the progression of metastasis of cancer cells and is a prognostic marker in several types of cancer (LI *et al.*, 2013).

The proteins THBS2, TIMP3, ACTA2 and IGF1 are all included in gene signatures of cancer (MATONDO *et al.*, 2017). The protein IGF1 has been associated with resistance of chemotherapy and it is a target molecule in the study of therapies for non-responders and partial remission patients (MATONDO *et al.*, 2017). IGF1 has another important role in the dynamic of the extra-cellular matrix. As PLAU, it can promote the invasiveness of the extra-cellular matrix (COX *et al.*, 2015). Cox *et al.* (2015) states that the functionality of IGF1 depends on features of the extra-cellular matrix and presence or absence of certain proteins.

The just presented medical findings about the proteins entering in our estimated confirmatory (and exploratory) models give support for understanding the validity of the configuration of model matrices used in the proposed models and also for interpreting the common factors in biological terms.

Figure 13 shows a diagram relating the 11 proteins entering our confirmatory model. Although the methods used for obtaining this diagram were completely distinct from the factor analysis model we used in our work, the final results were very alike. In our model and in the diagram it can be noted two clusters of highly correlated variables. Those variables group together to form the two common factors, validating the Kaiser criterion as effective for selecting the number of common latent factors in our applications. Although there are differences in the clusters of variables determined by the MCFA-SMN model and the diagram of the STRING software, those differences could be explained in terms of the medical findings described above. For example, the PLAU molecule has been described in laboratory experiments as being highly associated with molecules from the collagen linage. Based on this fact, our confirmatory model, although putting the PLAU molecule in a different cluster when compared with the STRING diagram, is still in accordance with the current scientific knowledge about the biological role of this molecule.

Finally, the interpretation of the two common latent factor entering in the confirmatory model can be explained as follows:

- the first common factor, characterized by the association of the proteins COL3A1, COL11A1, COL5A2, THBS2, PLAU, ACTA2 and TIMP3, could be interpreted as fundamentally associated with the production of collagen in the medium where cancer cells proliferate. This is justified since the higher loadings are those of COL3A1, COL11A1 and COL5A2, which are all collagen molecules.
- The second common factor determines the correlation between the molecules COL10A1, PDGFRA, PDGFRB and IGF1. Here, the higher loadings are those of COL10A1 and PDGFRB. According to the medical findings presented earlier in this section, the association of this two molecules could be pointing to the regulation of the density of the extracelular matrix and angiogenesis, *i.e* the proliferation of blood vessels (which are determinant in the metastasis of cancer).

It is important to notice that our confirmatory model gives a further information about the latent structure responsible for the association of the set of 11 targeted protein molecules. That piece of extra information is the correlation between common factors. In our confirmatory model the estimate of this correlation was above 0.9, reveling a strong association between the two biological functions described above in the interpretation of the common latent factors. The assumption of oblique common factors is supported also by the diagram of Figure 13.

As a conclusion, the proposed confirmatory model based on the results of our primary exploratory study is in line with experimental results obtained in medical research and also with the results generated from other approaches of data analysis applied to similar data sets.

5 Concluding remarks and further directions

5.1 Resumo da seção

Neste capítulo concluímos a dissertação retomando brevemente os principais pontos abordados na pesquisa e destacando a importância dos novos resultados obtidos. Também são feitas propostas para pesquisas futuras envolvendo o modelo MCFA-SMN, incluindo sua extensão para a classe de distribuições elípticas e para a análise de dados censurados.

5.2 Conclusions

In this dissertation, we have proposed a confirmatory factor analysis model that generalizes the model proposed by Jöreskog (1971), under SMN distribution for latent factors. We gave conditions for verifying parameter identification in the MCFA-SMN model through two simple theorems. Those theorems showed that all identification conditions for FA models for only one population can adequately be adapted to identify the MCFA-SMN model. The MCFA-SMN model defined and studied in this dissertation represents an important step towards the development of factor analysis models for simultaneous analysis of several populations. Our choice of algorithm for maximum likelihood estimation, the ECM algorithm, follows a trend in contemporary studies of FA models (CASTRO *et al.*, 2014; LIN *et al.*, 2014; ZHANG *et al.*, 2014; LIN *et al.*, 2016) and gives a simple framework for implementation and estimation of the MCFA-SMN model. Our simulation studies showed that the proposed estimators have good properties in finite samples. For the calculation of standard errors we have suggested two methods based on the literature about FA models (JAMSHIDIAN, 1997; LIN *et al.*, 2014). The simulation results showed that the method of Meilijson (1989) based on the empirical Fisher information matrix is preferable and leads to confidence intervals with probability coverage close to the nominal level.

Our application represents an important moment of our research since using the MCFA-SMN model we could extract meaningful interpretations from the estimated parameters, confirming the scientific knowledge developed in laboratory research in medicine and molecular biology. Hence, the MCFA-SMN model extends an important technique that can contribute to the development of other sciences, like Biology and

Medicine.

Although, further studies are necessary for confirming the good properties observed in our study for the MCFA-SMN model. Specially, larger simulation studies are necessary. New simulations studies could be designed to describe the behavior of estimators when more than two populations are considered in the model or when other choices of invariance structure for θ is assumed. Simulation studies directed to the evaluation of robust properties of the estimators are also important. This has been done in FA models for only one population while assuming latent factor following multivariate t-Student distribution (ZHANG *et al.*, 2014; CASTRO *et al.*, 2014; LAI and ZHANG, 2017).

Extensions of the MCFA-SMN model could also be a target for research. One simple extension is to consider an intercept in the model , with the intercept also dependent on θ. This new model would extend the model proposed by Sörbom (1974) and would allow to fit models with greater degree of invariance in θ. Also, the tobit-CFA model proposed by Castro *et al.* (2014) could be extended to the MCFA-SMN framework. It would lead to an MCFA-SMN model capable of dealing with censored observations. To extend the SMN class of distributions to the elliptical class of probability distributions (FANG and ZHANG, 1990) is also an important topic for research. Lemonte and Patriota (2011) have proposed a general class of multivariate elliptical models that, according to the authors, can account for the Structural Equation Models of Bollen (1989). In this case, the MCFA-SMN model would appear as a particular case of the model proposed by Lemonte and Patriota (2011). Although, Lemonte and Patriota (2011) does not discuss any of the association his model could have with factor analysis, *i.e.* the authors does not discuss in the FA context any kind of model specification or parameter identification. Although, Lemonte and Patriota (2011) gives a Newton-Raphson algorithm that could be used to estimate structural equation models with elliptically distributed latent factors. Hence, Lemonte and Patriota (2011) is an important starting point for further extensions of the MCFA-SMN model we discussed in this dissertation.

References

Adachi, K. Factor analysis with EM algorithm never gives improper solutions when sample covariance and initial parameter matrices are proper. *Psychometrika*, vol. 78(2):pp. 380–394, 2013.

Akaike, H. Factor analysis and AIC. *Psychometrika*, vol. 52(3):pp. 317–332, 1987.

Amatya, A.; Demirtas, H. Ordnor: An R package for concurrent generation of correlated ordinal and normal data. *Journal of Statistical Software, Code Snippets*, vol. 68(2):pp. 1–14, 2015.

Anderson, T. W.; Rubin, H. statistical inference in factor analysis. vol. 5, pp. 111–150. University of California Press, 1956.

Andrews, D. F.; Mallows, C. L. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 36(1):pp. 99–102, 1974.

Bai, J.; Li, K. Statistical analysis of factor models of high dimension. *The Annals of Statistics*, vol. 40(1):pp. 436–465, 2012.

Beale, E. M. L.; Mallows, C. L. Scale mixing of symmetric distributions with zero means. *The Annals of Mathematical Statistics*, vol. 30(4):pp. 1145–1151, 1959.

Bekker, P.; Merckens, A.; Wansbeek, T. *Identification, Equivalent Models, and Computer Algebra*. Statistical modeling and decision science. Academic Press, 1994.

Bekker, P. A.; ten Berge, J. M. Generic global indentification in factor analysis. *Linear Algebra and its Applications*, vol. 264:pp. 255–263, 1997.

Bentler, P. M. Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika*, vol. 48(4):pp. 493–517, 1983.

Bollen, K. A. *Structural Equations with Latent Variables*. John Wiley & Sons, Inc., 1989.

Bollen, K. A.; Jöreskog, K. G. Uniqueness does not imply identification. *Sociological Methods & Research*, vol. 14(2):pp. 155–163, 1985.

Brown, A. A.; *et al.*. Pathway-based factor analysis of gene expression data produces highly heritable phenotypes that associate with age. *G3: Genes, Genomes, Genetics*, vol. 5(5):pp. 839–847, 2015.

Browne, M. W.; Shapiro, A. Adjustments for kurtosis in factor analysis with elliptically distributed errors. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 49(3):pp. 346–352, 1987.

Buettner, F.; *et al.*. f-scLVM: scalable and versatile factor analysis for single-cell RNA-seq. *Genome Biology*, vol. 18(1):pp. 212–225, 2017.

Cabral-Marques, O.; Riemekasten, G. Functional autoantibodies targeting G protein-coupled receptors in rheumatic diseases. *Nature Reviews Rheumatology*, vol. 13(11):pp. 648–656, 2017.

Casey, T. M.; *et al.*. Cancer associated fibroblasts stimulated by transforming growth factor beta1 (TGF-beta 1) increase invasion rate of tumor cells: a population study. *Breast Cancer Research and Treatment*, vol. 110(1):pp. 39–49, 2007.

Castro, L. M.; *et al.*. Likelihood-based inference for tobit confirmatory factor analysis using the multivariate student-t distribution. *Statistics and Computing*, vol. 25(6):pp. 1163–1183, 2014.

Cox, O. T.; *et al.*. IGF-1 receptor and adhesion signaling: An important axis in determining cancer cell phenotype and therapy resistance. *Frontiers in Endocrinology*, vol. 6:p. 106, 2015.

De Vito, R. *Multi-study factor models for high-dimensional biological data*. Ph.D. thesis, Universita degli Studi di Padova, 2016.

Dempster, A. P.; Laird, N. M.; Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

Fang, K.-T.; Zhang, Y.-T. *Generalized Multivariate Analysis*. Springer, 1990.

Fang, M.; *et al.*. Collagen as a double-edged sword in tumor progression. *Tumor Biology*, vol. 35(4):pp. 2871–2882, 2014.

Ferreira, C. S.; Lachos, V. H.; Bolfarine, H. Likelihood-based inference for multivariate skew scale mixtures of normal distributions. *AStA Advances in Statistical Analysis*, vol. 100(4):pp. 421–441, 2016.

Gascard, P.; Tlsty, T. D. Carcinoma-associated fibroblasts: orchestrating the composition of malignancy. *Genes & Development*, vol. 30(9):pp. 1002–1019, 2016.

Genser, B.; *et al.*. A guide to modern statistical analysis of immunological data. *BMC Immunology*, vol. 8(1):pp. 27–42, 2007.

Germain, R. N.; *et al.*. Systems biology in immunology: A computational modeling perspective. *Annual Review of Immunology*, vol. 29(1):pp. 527–585, 2011.

Geweke, J.; Zhou, G. Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies*, vol. 9(2):pp. 557–587, 1996.

Gialeli, C.; *et al.*. PDGF/PDGFR signaling and targeting in cancer growth and progression: Focus on tumor microenvironment and cancer-associated fibroblasts. *Current Pharmaceutical Design*, vol. 20(17):pp. 2843–2848, 2014.

Hammer, A. M.; *et al.*. Stromal PDGFR-α activation enhances matrix stiffness, impedes mammary ductal development, and accelerates tumor growth. *Neoplasia*, vol. 19(6):pp. 496–508, 2017.

Hayashi, K.; Liang, L. On the detection of improper solutions in factor analayis by the EM algorithm. *Behaviormetrika*, vol. 41(2):pp. 245–268, 2014.

Hoffmann, G. W. A neural network model based on the analogy with the immune system. *Journal of Theoretical Biology*, vol. 122(1):pp. 33–67, 1986.

Holzinger, K. J.; Swineford, F. *A study in factor analysis: The stability of a bi-factor solution*. Ph.D. thesis, University of Chicago, 1939.

Iacob, E.; *et al.*. Gene expression factor analysis to differentiate pathways linked to fibromyalgia, chronic fatigue syndrome, and depression in a diverse patient sample. *Arthritis Care & Research*, vol. 68(1):pp. 132–140, 2016.

Jamshidian, M. *Latent Variable Modeling and Applications to Causality*, chap. 13, pp. 247–258. Springer Verlag, 1997.

Jamshidian, M.; Jennrich, R. I. Conjugate gradient methods in confirmatory factor analysis. *Computational Statistics & Data Analysis*, vol. 17(3):pp. 247–263, 1994.

Jia, D.; *et al.*. A COL11A1-correlated pan-cancer gene signature of activated fibroblasts for the prioritization of therapeutic targets. *Cancer Letters*, vol. 382(2):pp. 203–214, 2016.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. *ETS Research Bulletin Series*, vol. 32(2):pp. 443–482, 1967.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *ETS Research Bulletin Series*, vol. 34(2):pp. 183–202, 1969.

Jöreskog, K. G. Simultaneous factor analysis in several populations. *Psychometrika*, vol. 36(4):pp. 409–426, 1971.

Kano, Y. Consistency property of elliptic probability density functions. *Journal of Multivariate Analysis*, vol. 51(1):pp. 139–147, 1994.

Kaufman, J. D.; Dunlap, W. P. Determining the number of factors to retain: Q windows-based FORTRAN-IMSL program for parallel analysis. *Behavior Research Methods, Instruments, & Computers*, vol. 32(3):pp. 389–395, 2000.

Kerr, M. K. Experimental design for gene expression microarrays. *Biostatistics*, vol. 2(2):pp. 183–201, 2001.

Lai, M. H. C.; Zhang, J. Evaluating fit indices for multivariate t-based structural equation modeling with data contamination. *Frontiers in Psychology*, vol. 8:pp. 1–13, 2017.

Lemonte, A. J.; Patriota, A. G. Multivariate elliptical models with general parameterization. *Statistical Methodology*, vol. 8(4):pp. 389–400, 2011.

Li, D.; *et al.*. The critical role of dysregulated FOXM1-PLAUR signaling in human colon cancer progression and metastasis. *Clinical Cancer Research*, vol. 19(1):pp. 62–72, 2013.

Lin, T.-I.; McLachlan, G. J.; Lee, S. X. Extending mixtures of factor models using the restricted multivariate skew-normal distribution. *Journal of Multivariate Analysis*, vol. 143:pp. 398–413, 2016.

Lin, T.-I.; *et al.*. A robust factor analysis model using the restricted skew-t distribution. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, vol. 24(3):pp. 510–531, 2014.

Magnus, J. R. On the concept of matrix derivative. *Journal of Multivariate Analysis*, vol. 101(9):pp. 2200–2206, 2010.

Magnus, J. R.; Neudecker, H. Matrix differential calculus with applications to simple, hadamard, and kronecker products. *Journal of Mathematical Psychology*, vol. 29(4):pp. 474–492, 1985.

Matondo, A.; *et al.*. The prognostic 97 chemoresponse gene signature in ovarian cancer. *Scientific Reports*, vol. 7(1):pp. 1–12, 2017.

Mehra, N. Standardized versus unstandardized factor analysis in a study of "organizational climate". *The Journal of Experimental Education*, vol. 42(2):pp. 60–67, 1973.

Meilijson, I. A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 51(1):pp. 127–138, 1989.

Meng, X.-L.; Rubin, D. B. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, vol. 80(2):pp. 267–278, 1993.

Meredith, W. Rotation to achieve factorial invariance. *Psychometrika*, vol. 29(2):pp. 187–206, 1964.

Meredith, W.; Teresi, J. A. An essay on measurement and factorial invariance. *Medical Care*, vol. 44(11):pp. S69–S77, 2006.

Montanari, A.; Viroli, C. Heteroscedastic factor mixture analysis. *Statistical Modelling: An International Journal*, vol. 10(4):pp. 441–460, 2010.

Murphy, J. R. Statistical errors in immunologic research. *Journal of Allergy and Clinical Immunology*, vol. 114(6):pp. 1259–1263, 2004.

Myers, N. D.; *et al.*. Reordering and reflecting factors for simulation studies with exploratory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 24(1):pp. 112–128, 2016.

Peeters, C. F. W. Rotational uniqueness conditions under oblique factor correlation metric. *Psychometrika*, vol. 77(2):pp. 288–292, 2012.

Perelson, A. S. Immune network theory. *Immunological Reviews*, vol. 110(1):pp. 5–36, 1989.

Pham, L. M.; *et al.*. Perturbation detection through modeling of gene expression on a latent biological pathway network: A bayesian hierarchical approach. *Journal of the American Statistical Association*, vol. 111(513):pp. 73–92, 2016.

Ponder, B. A. J. Cancer genetics. *Nature*, vol. 411(6835):pp. 336–341, 2001.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

Reilly, T. A necessary and sufficient condition for identification of confirmatory factor analysis models of factor complexity one. *Sociological Methods & Research*, vol. 23(4):pp. 421–441, 1995.

Reilly, T.; O'Brien, R. M. Identification of confirmatory factor analysis models of arbitrary complexity. *Sociological Methods & Research*, vol. 24(4):pp. 473–491, 1996.

Rieckmann, J. C.; *et al.*. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nature Immunology*, vol. 18(5):pp. 583–593, 2017.

Rosseel, Y. lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, vol. 48(2):pp. 1–36, 2012.

Roy, S.; Bhattacharyya, D. K.; Kalita, J. K. Reconstruction of gene co-expression network from microarray data using local expression patterns. *BMC Bioinformatics*, vol. 15(7):p. S10, 2014.

Rubin, D. B.; Thayer, D. T. EM algorithms for ML factor analysis. *Psychometrika*, vol. 47(1):pp. 69–76, 1982.

Song, X.-Y.; Lee, S.-Y. Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical and Statistical Psychology*, vol. 54(2):pp. 237–263, 2001.

Sörbom, D. A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, vol. 27(2):pp. 229–239, 1974.

Teng, P.-N.; *et al.*. Identification of candidate circulating cisplatin-resistant biomarkers from epithelial ovarian carcinoma cell secretomes. *British Journal of Cancer*, vol. 110(1):pp. 123–132, 2013.

Wang, W.-L.; Lin, T.-I.; Lachos, V. H. Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Statistical Methods in Medical Research*, vol. 27(1):pp. 48–64, 2015.

Wang, X. V.; Parmigiani, G. Integrative factor analysis — an unsupervised method for quantifying cross-study consistency of gene expression data. *Genomics*, vol. 110(2):pp. 80–88, 2018.

West, M. Outlier models and prior distributions in bayesian linear regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 46(3):pp. 431–439, 1984.

West, M. On scale mixtures of normal distributions. *Biometrika*, vol. 74(3):pp. 646–648, 1987.

Yuan, K.-H.; Bentler, P. M. On chi-square difference and z tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement*, vol. 64(5):pp. 737–757, 2004.

Yuan, K.-H.; Bentler, P. M. Mean comparison: Manifest variable versus latent variable. *Psychometrika*, vol. 71(1):pp. 139–159, 2006.

Yuan, K.-H.; Chan, W. Measurement invariance via multigroup SEM: Issues and solutions with chi-square-difference tests. *Psychological Methods*, vol. 21(3):pp. 405–426, 2016.

Zhang, J.; Li, J.; Liu, C. Robust factor analysis using the multivariate t-distribution. *Statistica Sinica*, vol. 24:pp. 291–312, 2014.

Zhong, X.; Yuan, K.-H. Bias and efficiency in structural equation modeling: Maximum likelihood versus robust methods. *Multivariate Behavioral Research*, vol. 46(2):pp. 229–265, 2011.

Appendix A - Matrix calculus

The proposed ECM algorithm for estimation of our new model relies on differentiation of matrices. Matrix derivatives have several different definitions in the literature, related mainly to notation and the way partial derivatives are organized in a new matrix (MAGNUS and NEUDECKER, 1985). Magnus (2010) extended the differentiation rules of vector calculus to the framework of matrix calculus. We adopt his definition, which is stated below.

Definition 4. *Let F be an $m \times p$ matrix function of a matrix of variables $\boldsymbol{X}$ with dimension $n \times q$. The derivative of F with respect to X is defined as the $mp \times nq$ matrix*

$$DF(\boldsymbol{X}) = \frac{\partial vec(F(\boldsymbol{X}))}{\partial vec(\boldsymbol{X})^\top}.$$

Hence, the rules of differentiation of matrices enjoys the same nice properties of vector calculus, the most important being the chain rule. Magnus and Neudecker (1985) formally states the chain rule of matrix calculus. We shall state it loosely, following Magnus (2010).

Definition 5. *Let $\boldsymbol{X}$ be a $n \times q$ matrix of variables, F, $m \times p$, differentiable at $\boldsymbol{X}$ and G, $l \times r$ diferentiable at $\boldsymbol{Y} = F(\boldsymbol{X})$. Then, $H(\boldsymbol{X}) = G(F(\boldsymbol{X}))$ is differentiable at $\boldsymbol{X}$, and*

$$DH(\boldsymbol{X}) = \frac{\partial vec(G(\boldsymbol{Y}))}{\partial vec(\boldsymbol{Y})^\top} \frac{\partial vec(F(\boldsymbol{X}))}{\partial vec(\boldsymbol{X})^\top}.$$

The notation vec adopted in Definitions (4) and (5) refers to the vec operator, which stacks the columns of a matrix $\boldsymbol{A}$, $p \times q$, one beneath the other to get a unique pq-dimensional column vector vec$(\boldsymbol{A})$. Analogously, the vech operator puts $\boldsymbol{A} = (a_{ij})$ in a vector form, but only taking the elements a_{ij} whose $i \geq j$. The diag operator extracts the diagonal of $\boldsymbol{A}$ and present it as a column vector. Next we define the duplication matrix, $\boldsymbol{D}$, and the *diag* matrix, $\boldsymbol{B}$.

Definition 6. *Let $\boldsymbol{A}$ be a square matrix of order p. The duplication matrix $\boldsymbol{D}$ has dimensions $p^2 \times p(p+1)/2$, and is implicitly defined as*

$$\boldsymbol{D} vech(\boldsymbol{A}) = vec(\boldsymbol{A}). \tag{A.1}$$

Definition 7. *Let $\boldsymbol{A}$ be a diagonal matrix of order p. The diag matrix $\boldsymbol{B}$ has dimensions $p^2 \times p$, and is implicitly defined as*

$$\boldsymbol{B}\,\textit{diag}(\boldsymbol{A}) = \textit{vec}(\boldsymbol{A}). \tag{A.2}$$

Appendix B - Tables of simulations' results

Here, we give tables with the simulation results discussed in Section 3.

Table 10: Bias.

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
1	λ_{21}	0.0021	0.0022	0.0049	0.0018	200
2	λ_{21}	0.0015	0.0007	0.0023	0.0013	400
3	λ_{21}	0.0011	0.0025	0.0005	-0.0002	600
4	λ_{21}	0.0005	0.0015	0.0001	-0.0001	800
5	λ_{21}	0.0013	0.0011	0.0015	0.0013	1000
6	λ_{31}	0.0020	0.0077	0.0054	0.0035	200
7	λ_{31}	0.0017	0.0030	0.0049	0.0038	400
8	λ_{31}	0.0013	0.0042	0.0025	0.0006	600
9	λ_{31}	0.0020	0.0018	0.0014	0.0008	800
10	λ_{31}	0.0014	0.0026	0.0015	0.0027	1000
11	λ_{52}	0.0014	-0.0883	0.0020	0.0023	200
12	λ_{52}	0.0011	0.0016	0.0008	0.0004	400
13	λ_{52}	0.0005	0.0004	0.0010	0.0008	600
14	λ_{52}	0.0007	0.0004	-0.0004	0.0004	800
15	λ_{52}	-0.0003	0.0008	0.0010	0.0001	1000
16	λ_{62}	0.0014	-0.0781	0.0012	0.0020	200
17	λ_{62}	0.0009	0.0012	0.0009	0.0011	400
18	λ_{62}	0.0002	0.0010	0.0012	0.0012	600
19	λ_{62}	0.0008	0.0007	0.0004	0.0002	800
20	λ_{62}	-0.0001	0.0011	0.0005	0.0013	1000
21	λ_{83}	-0.0387	-0.0162	0.0146	-0.0503	200
22	λ_{83}	0.0056	0.0056	0.0058	0.0052	400
23	λ_{83}	0.0012	0.0007	0.0052	0.0046	600
24	λ_{83}	0.0041	0.0050	0.0027	0.0052	800
25	λ_{83}	0.0031	0.0016	0.0023	0.0027	1000
26	λ_{93}	-0.0371	-0.0096	0.0115	-0.0496	200
27	λ_{93}	0.0053	0.0064	0.0040	0.0025	400
28	λ_{93}	0.0026	0.0004	0.0034	0.0029	600
29	λ_{93}	0.0033	0.0037	0.0018	0.0046	800
30	λ_{93}	0.0027	0.0014	0.0023	0.0015	1000
31	ψ_{11}	-0.0058	-0.0032	-0.0031	-0.0043	200
32	ψ_{11}	-0.0014	-0.0021	-0.0007	-0.0017	400
33	ψ_{11}	-0.0017	0.0007	-0.0013	-0.0010	600
34	ψ_{11}	-0.0006	0.0000	-0.0007	-0.0011	800
35	ψ_{11}	-0.0008	-0.0005	-0.0004	-0.0007	1000
36	ψ_{22}	-0.0022	-0.0006	-0.0018	-0.0012	200

(continue in the next page)

Table 10: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
37	ψ_{22}	-0.0018	0.0003	-0.0011	-0.0001	400
38	ψ_{22}	-0.0004	0.0002	-0.0017	-0.0004	600
39	ψ_{22}	-0.0013	-0.0000	-0.0012	0.0002	800
40	ψ_{22}	-0.0003	0.0006	-0.0013	-0.0004	1000
41	ψ_{33}	-0.0016	-0.0029	-0.0029	-0.0010	200
42	ψ_{33}	-0.0005	-0.0018	-0.0013	-0.0009	400
43	ψ_{33}	-0.0002	-0.0008	-0.0011	-0.0006	600
44	ψ_{33}	-0.0006	-0.0004	-0.0007	-0.0008	800
45	ψ_{33}	-0.0012	-0.0013	-0.0005	-0.0013	1000
46	ψ_{44}	-0.0012	0.0004	-0.0002	-0.0001	200
47	ψ_{44}	-0.0004	0.0002	-0.0001	-0.0003	400
48	ψ_{44}	-0.0003	0.0000	-0.0006	0.0000	600
49	ψ_{44}	-0.0003	-0.0002	-0.0006	-0.0007	800
50	ψ_{44}	-0.0005	-0.0001	0.0000	-0.0002	1000
51	ψ_{55}	-0.0008	-0.0005	-0.0016	-0.0007	200
52	ψ_{55}	-0.0004	-0.0003	-0.0005	-0.0005	400
53	ψ_{55}	-0.0007	0.0002	-0.0005	-0.0003	600
54	ψ_{55}	-0.0002	-0.0000	0.0001	0.0003	800
55	ψ_{55}	-0.0003	0.0003	-0.0004	0.0000	1000
56	ψ_{66}	-0.0011	-0.0011	-0.0009	-0.0013	200
57	ψ_{66}	-0.0006	0.0001	-0.0006	-0.0005	400
58	ψ_{66}	-0.0001	0.0003	-0.0005	-0.0002	600
59	ψ_{66}	-0.0002	-0.0001	-0.0004	0.0002	800
60	ψ_{66}	0.0000	-0.0004	-0.0001	-0.0005	1000
61	ψ_{77}	-0.0025	-0.0014	-0.0017	-0.0015	200
62	ψ_{77}	-0.0005	-0.0002	-0.0011	-0.0014	400
63	ψ_{77}	-0.0012	-0.0015	-0.0006	-0.0002	600
64	ψ_{77}	-0.0006	0.0002	-0.0013	-0.0000	800
65	ψ_{77}	-0.0003	-0.0003	-0.0004	-0.0002	1000
66	ψ_{88}	-0.0039	-0.0021	-0.0034	-0.0034	200
67	ψ_{88}	-0.0001	-0.0008	-0.0027	-0.0005	400
68	ψ_{88}	-0.0000	-0.0003	-0.0018	-0.0010	600
69	ψ_{88}	-0.0008	-0.0007	-0.0012	-0.0009	800
70	ψ_{88}	-0.0008	-0.0005	-0.0002	-0.0018	1000
71	ψ_{99}	-0.0031	-0.0007	-0.0024	-0.0047	200
72	ψ_{99}	-0.0018	-0.0022	-0.0010	-0.0012	400
73	ψ_{99}	-0.0016	-0.0007	-0.0010	-0.0009	600
74	ψ_{99}	-0.0012	-0.0005	-0.0002	-0.0005	800
75	ψ_{99}	-0.0006	-0.0005	-0.0007	0.0004	1000

(continue in the next page)

Table 10: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
76	ζ_{11}	0.0053	0.0048	0.0019	0.0070	200
77	ζ_{11}	0.0015	0.0033	0.0002	0.0011	400
78	ζ_{11}	0.0020	0.0003	0.0012	0.0025	600
79	ζ_{11}	0.0001	0.0010	0.0011	0.0021	800
80	ζ_{11}	0.0006	0.0013	-0.0001	0.0003	1000
81	ζ_{21}	-0.0012	0.0001	0.0000	0.0018	200
82	ζ_{21}	-0.0003	0.0002	-0.0006	-0.0000	400
83	ζ_{21}	0.0006	-0.0000	-0.0001	0.0005	600
84	ζ_{21}	0.0001	-0.0004	-0.0000	0.0001	800
85	ζ_{21}	0.0005	0.0002	-0.0002	-0.0002	1000
86	ζ_{22}	0.0005	0.0025	0.0009	0.0020	200
87	ζ_{22}	0.0006	0.0001	0.0004	0.0004	400
88	ζ_{22}	0.0009	0.0013	0.0003	0.0004	600
89	ζ_{22}	0.0003	0.0006	0.0006	0.0006	800
90	ζ_{22}	0.0016	0.0003	0.0002	0.0002	1000
91	ζ_{31}	-0.0003	0.0009	-0.0003	0.0008	200
92	ζ_{31}	0.0001	0.0004	-0.0005	0.0003	400
93	ζ_{31}	0.0007	0.0004	0.0000	0.0004	600
94	ζ_{31}	-0.0001	0.0004	0.0003	0.0001	800
95	ζ_{31}	0.0001	0.0003	-0.0000	-0.0002	1000
96	ζ_{32}	0.0000	0.0006	-0.0004	0.0007	200
97	ζ_{32}	-0.0002	0.0001	0.0002	0.0002	400
98	ζ_{32}	0.0004	0.0005	-0.0001	0.0001	600
99	ζ_{32}	-0.0000	0.0001	0.0001	-0.0001	800
100	ζ_{32}	-0.0001	0.0003	0.0000	-0.0001	1000
101	ζ_{33}	0.0029	0.0056	0.0022	0.0044	200
102	ζ_{33}	0.0012	0.0032	0.0019	0.0019	400
103	ζ_{33}	0.0015	0.0030	0.0010	0.0012	600
104	ζ_{33}	-0.0000	0.0009	0.0008	-0.0001	800
105	ζ_{33}	0.0005	0.0015	0.0006	0.0008	1000

Table 11: Mean Square Error (MSE).

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
1	λ_{21}	0.0081	0.0095	0.0088	0.0087	200
2	λ_{21}	0.0040	0.0046	0.0042	0.0043	400
3	λ_{21}	0.0026	0.0032	0.0028	0.0028	600
4	λ_{21}	0.0020	0.0023	0.0021	0.0022	800
5	λ_{21}	0.0016	0.0019	0.0017	0.0017	1000
6	λ_{31}	0.0112	0.0130	0.0115	0.0116	200

(continue in the next page)

Table 11: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
7	λ_{31}	0.0053	0.0064	0.0058	0.0057	400
8	λ_{31}	0.0035	0.0042	0.0039	0.0037	600
9	λ_{31}	0.0027	0.0032	0.0028	0.0028	800
10	λ_{31}	0.0022	0.0024	0.0023	0.0023	1000
11	λ_{52}	0.0030	14.0492	0.0032	0.0031	200
12	λ_{52}	0.0015	0.0017	0.0015	0.0015	400
13	λ_{52}	0.0009	0.0011	0.0010	0.0010	600
14	λ_{52}	0.0007	0.0008	0.0008	0.0008	800
15	λ_{52}	0.0006	0.0007	0.0006	0.0006	1000
16	λ_{62}	0.0029	11.3490	0.0031	0.0031	200
17	λ_{62}	0.0014	0.0017	0.0015	0.0015	400
18	λ_{62}	0.0010	0.0011	0.0011	0.0010	600
19	λ_{62}	0.0007	0.0008	0.0008	0.0008	800
20	λ_{62}	0.0006	0.0007	0.0006	0.0006	1000
21	λ_{83}	3.1571	3.5002	0.0248	3.7722	200
22	λ_{83}	0.0112	0.0130	0.0116	0.0115	400
23	λ_{83}	0.0072	0.0084	0.0077	0.0075	600
24	λ_{83}	0.0055	0.0062	0.0058	0.0057	800
25	λ_{83}	0.0043	0.0050	0.0046	0.0045	1000
26	λ_{93}	2.9575	2.3612	0.0201	3.6304	200
27	λ_{93}	0.0091	0.0107	0.0095	0.0095	400
28	λ_{93}	0.0057	0.0066	0.0063	0.0063	600
29	λ_{93}	0.0044	0.0051	0.0047	0.0047	800
30	λ_{93}	0.0035	0.0041	0.0038	0.0037	1000
31	ψ_{11}	0.0051	0.0057	0.0047	0.0049	200
32	ψ_{11}	0.0022	0.0027	0.0024	0.0023	400
33	ψ_{11}	0.0015	0.0018	0.0015	0.0016	600
34	ψ_{11}	0.0011	0.0014	0.0011	0.0012	800
35	ψ_{11}	0.0009	0.0011	0.0010	0.0009	1000
36	ψ_{22}	0.0040	0.0054	0.0044	0.0042	200
37	ψ_{22}	0.0019	0.0027	0.0022	0.0021	400
38	ψ_{22}	0.0014	0.0017	0.0015	0.0015	600
39	ψ_{22}	0.0010	0.0014	0.0011	0.0011	800
40	ψ_{22}	0.0008	0.0011	0.0009	0.0009	1000
41	ψ_{33}	0.0038	0.0050	0.0041	0.0040	200
42	ψ_{33}	0.0019	0.0024	0.0020	0.0020	400
43	ψ_{33}	0.0012	0.0016	0.0013	0.0013	600
44	ψ_{33}	0.0009	0.0012	0.0010	0.0010	800
45	ψ_{33}	0.0007	0.0009	0.0008	0.0008	1000

(continue in the next page)

Table 11: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
46	ψ_{44}	0.0010	0.0015	0.0011	0.0011	200
47	ψ_{44}	0.0005	0.0006	0.0005	0.0005	400
48	ψ_{44}	0.0003	0.0004	0.0004	0.0003	600
49	ψ_{44}	0.0003	0.0003	0.0003	0.0003	800
50	ψ_{44}	0.0002	0.0003	0.0002	0.0002	1000
51	ψ_{55}	0.0010	0.0013	0.0011	0.0011	200
52	ψ_{55}	0.0005	0.0007	0.0006	0.0005	400
53	ψ_{55}	0.0003	0.0004	0.0004	0.0004	600
54	ψ_{55}	0.0002	0.0003	0.0003	0.0003	800
55	ψ_{55}	0.0002	0.0003	0.0002	0.0002	1000
56	ψ_{66}	0.0011	0.0014	0.0011	0.0011	200
57	ψ_{66}	0.0005	0.0007	0.0006	0.0006	400
58	ψ_{66}	0.0003	0.0004	0.0004	0.0004	600
59	ψ_{66}	0.0003	0.0003	0.0003	0.0003	800
60	ψ_{66}	0.0002	0.0003	0.0002	0.0002	1000
61	ψ_{77}	0.0036	0.0048	0.0040	0.0039	200
62	ψ_{77}	0.0018	0.0024	0.0019	0.0019	400
63	ψ_{77}	0.0012	0.0015	0.0013	0.0013	600
64	ψ_{77}	0.0009	0.0012	0.0010	0.0009	800
65	ψ_{77}	0.0007	0.0009	0.0008	0.0008	1000
66	ψ_{88}	0.0040	0.0048	0.0043	0.0043	200
67	ψ_{88}	0.0020	0.0024	0.0021	0.0021	400
68	ψ_{88}	0.0013	0.0016	0.0014	0.0014	600
69	ψ_{88}	0.0010	0.0011	0.0010	0.0010	800
70	ψ_{88}	0.0008	0.0010	0.0008	0.0008	1000
71	ψ_{99}	0.0036	0.0046	0.0039	0.0038	200
72	ψ_{99}	0.0018	0.0022	0.0019	0.0019	400
73	ψ_{99}	0.0012	0.0014	0.0013	0.0012	600
74	ψ_{99}	0.0009	0.0012	0.0010	0.0010	800
75	ψ_{99}	0.0007	0.0009	0.0007	0.0008	1000
76	ζ_{11}	0.0078	0.0095	0.0080	0.0082	200
77	ζ_{11}	0.0036	0.0047	0.0040	0.0039	400
78	ζ_{11}	0.0025	0.0030	0.0026	0.0026	600
79	ζ_{11}	0.0018	0.0023	0.0019	0.0020	800
80	ζ_{11}	0.0015	0.0018	0.0016	0.0016	1000
81	ζ_{21}	0.0021	0.0027	0.0023	0.0023	200
82	ζ_{21}	0.0011	0.0013	0.0012	0.0011	400
83	ζ_{21}	0.0007	0.0009	0.0008	0.0008	600
84	ζ_{21}	0.0005	0.0007	0.0006	0.0006	800

(continue in the next page)

Table 11: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
85	ζ_{21}	0.0004	0.0005	0.0005	0.0005	1000
86	ζ_{22}	0.0049	0.0068	0.0056	0.0055	200
87	ζ_{22}	0.0025	0.0033	0.0029	0.0027	400
88	ζ_{22}	0.0017	0.0022	0.0019	0.0018	600
89	ζ_{22}	0.0013	0.0017	0.0015	0.0014	800
90	ζ_{22}	0.0010	0.0014	0.0011	0.0011	1000
91	ζ_{31}	0.0015	0.0018	0.0015	0.0016	200
92	ζ_{31}	0.0007	0.0009	0.0008	0.0007	400
93	ζ_{31}	0.0005	0.0006	0.0005	0.0005	600
94	ζ_{31}	0.0004	0.0004	0.0004	0.0004	800
95	ζ_{31}	0.0003	0.0003	0.0003	0.0003	1000
96	ζ_{32}	0.0013	0.0015	0.0014	0.0014	200
97	ζ_{32}	0.0006	0.0008	0.0007	0.0007	400
98	ζ_{32}	0.0004	0.0005	0.0005	0.0004	600
99	ζ_{32}	0.0003	0.0004	0.0003	0.0003	800
100	ζ_{32}	0.0003	0.0003	0.0003	0.0003	1000
101	ζ_{33}	0.0043	0.0052	0.0045	0.0046	200
102	ζ_{33}	0.0022	0.0026	0.0022	0.0022	400
103	ζ_{33}	0.0014	0.0017	0.0015	0.0015	600
104	ζ_{33}	0.0010	0.0013	0.0011	0.0011	800
105	ζ_{33}	0.0009	0.0010	0.0009	0.0009	1000

Table 12: Monte Carlo standard errors (MCSE).

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
1	λ_{21}	0.0899	0.0977	0.0938	0.0930	200
2	λ_{21}	0.0633	0.0681	0.0651	0.0653	400
3	λ_{21}	0.0514	0.0567	0.0526	0.0524	600
4	λ_{21}	0.0449	0.0476	0.0453	0.0464	800
5	λ_{21}	0.0396	0.0433	0.0416	0.0409	1000
6	λ_{31}	0.1059	0.1137	0.1072	0.1078	200
7	λ_{31}	0.0730	0.0799	0.0763	0.0755	400
8	λ_{31}	0.0594	0.0646	0.0623	0.0612	600
9	λ_{31}	0.0518	0.0563	0.0532	0.0528	800
10	λ_{31}	0.0464	0.0491	0.0484	0.0477	1000
11	λ_{52}	0.0544	3.7472	0.0565	0.0559	200
12	λ_{52}	0.0386	0.0412	0.0391	0.0393	400
13	λ_{52}	0.0304	0.0338	0.0319	0.0313	600
14	λ_{52}	0.0266	0.0291	0.0280	0.0277	800

(continue in the next page)

Table 12: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
15	λ_{52}	0.0238	0.0261	0.0248	0.0249	1000
16	λ_{62}	0.0540	3.3679	0.0560	0.0554	200
17	λ_{62}	0.0377	0.0414	0.0393	0.0387	400
18	λ_{62}	0.0311	0.0338	0.0326	0.0317	600
19	λ_{62}	0.0268	0.0289	0.0280	0.0276	800
20	λ_{62}	0.0238	0.0260	0.0244	0.0252	1000
21	λ_{83}	1.7764	1.8708	0.1567	1.9416	200
22	λ_{83}	0.1057	0.1140	0.1075	0.1072	400
23	λ_{83}	0.0846	0.0915	0.0878	0.0867	600
24	λ_{83}	0.0742	0.0789	0.0760	0.0755	800
25	λ_{83}	0.0659	0.0705	0.0677	0.0671	1000
26	λ_{93}	1.7193	1.5366	0.1414	1.9047	200
27	λ_{93}	0.0951	0.1034	0.0976	0.0977	400
28	λ_{93}	0.0755	0.0815	0.0792	0.0793	600
29	λ_{93}	0.0666	0.0715	0.0688	0.0684	800
30	λ_{93}	0.0593	0.0639	0.0618	0.0610	1000
31	ψ_{11}	0.0711	0.0753	0.0685	0.0700	200
32	ψ_{11}	0.0468	0.0522	0.0487	0.0482	400
33	ψ_{11}	0.0388	0.0426	0.0388	0.0394	600
34	ψ_{11}	0.0333	0.0368	0.0339	0.0345	800
35	ψ_{11}	0.0295	0.0325	0.0309	0.0307	1000
36	ψ_{22}	0.0634	0.0735	0.0661	0.0647	200
37	ψ_{22}	0.0441	0.0524	0.0466	0.0460	400
38	ψ_{22}	0.0370	0.0418	0.0383	0.0382	600
39	ψ_{22}	0.0315	0.0375	0.0335	0.0326	800
40	ψ_{22}	0.0280	0.0329	0.0298	0.0295	1000
41	ψ_{33}	0.0617	0.0705	0.0641	0.0635	200
42	ψ_{33}	0.0434	0.0493	0.0449	0.0448	400
43	ψ_{33}	0.0350	0.0396	0.0367	0.0364	600
44	ψ_{33}	0.0306	0.0343	0.0319	0.0314	800
45	ψ_{33}	0.0270	0.0306	0.0290	0.0276	1000
46	ψ_{44}	0.0322	0.0392	0.0333	0.0328	200
47	ψ_{44}	0.0229	0.0255	0.0234	0.0232	400
48	ψ_{44}	0.0182	0.0209	0.0195	0.0185	600
49	ψ_{44}	0.0159	0.0180	0.0167	0.0167	800
50	ψ_{44}	0.0142	0.0160	0.0149	0.0149	1000
51	ψ_{55}	0.0321	0.0364	0.0332	0.0328	200
52	ψ_{55}	0.0226	0.0255	0.0235	0.0229	400
53	ψ_{55}	0.0182	0.0206	0.0193	0.0188	600

(continue in the next page)

Table 12: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
54	ψ_{55}	0.0156	0.0178	0.0167	0.0165	800
55	ψ_{55}	0.0142	0.0160	0.0148	0.0147	1000
56	ψ_{66}	0.0328	0.0368	0.0338	0.0335	200
57	ψ_{66}	0.0228	0.0259	0.0240	0.0237	400
58	ψ_{66}	0.0186	0.0210	0.0193	0.0194	600
59	ψ_{66}	0.0165	0.0183	0.0170	0.0168	800
60	ψ_{66}	0.0144	0.0161	0.0148	0.0150	1000
61	ψ_{77}	0.0596	0.0694	0.0629	0.0621	200
62	ψ_{77}	0.0424	0.0486	0.0433	0.0435	400
63	ψ_{77}	0.0342	0.0386	0.0358	0.0359	600
64	ψ_{77}	0.0295	0.0342	0.0311	0.0305	800
65	ψ_{77}	0.0263	0.0300	0.0277	0.0276	1000
66	ψ_{88}	0.0627	0.0694	0.0652	0.0653	200
67	ψ_{88}	0.0447	0.0492	0.0459	0.0459	400
68	ψ_{88}	0.0357	0.0406	0.0373	0.0372	600
69	ψ_{88}	0.0309	0.0339	0.0322	0.0319	800
70	ψ_{88}	0.0279	0.0309	0.0288	0.0285	1000
71	ψ_{99}	0.0599	0.0676	0.0624	0.0613	200
72	ψ_{99}	0.0423	0.0469	0.0436	0.0430	400
73	ψ_{99}	0.0346	0.0377	0.0359	0.0349	600
74	ψ_{99}	0.0298	0.0340	0.0315	0.0308	800
75	ψ_{99}	0.0264	0.0302	0.0272	0.0276	1000
76	ζ_{11}	0.0881	0.0974	0.0894	0.0900	200
77	ζ_{11}	0.0604	0.0683	0.0632	0.0622	400
78	ζ_{11}	0.0502	0.0545	0.0513	0.0513	600
79	ζ_{11}	0.0427	0.0475	0.0437	0.0450	800
80	ζ_{11}	0.0389	0.0421	0.0404	0.0397	1000
81	ζ_{21}	0.0461	0.0522	0.0483	0.0476	200
82	ζ_{21}	0.0330	0.0363	0.0346	0.0334	400
83	ζ_{21}	0.0269	0.0300	0.0278	0.0276	600
84	ζ_{21}	0.0230	0.0256	0.0239	0.0241	800
85	ζ_{21}	0.0212	0.0228	0.0218	0.0213	1000
86	ζ_{22}	0.0703	0.0826	0.0750	0.0741	200
87	ζ_{22}	0.0499	0.0576	0.0535	0.0522	400
88	ζ_{22}	0.0412	0.0468	0.0434	0.0427	600
89	ζ_{22}	0.0360	0.0409	0.0382	0.0373	800
90	ζ_{22}	0.0318	0.0370	0.0336	0.0334	1000
91	ζ_{31}	0.0381	0.0420	0.0388	0.0399	200
92	ζ_{31}	0.0269	0.0294	0.0279	0.0273	400

(continue in the next page)

Table 12: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
93	ζ_{31}	0.0220	0.0240	0.0228	0.0226	600
94	ζ_{31}	0.0188	0.0205	0.0193	0.0196	800
95	ζ_{31}	0.0173	0.0184	0.0174	0.0176	1000
96	ζ_{32}	0.0362	0.0392	0.0369	0.0371	200
97	ζ_{32}	0.0252	0.0276	0.0266	0.0256	400
98	ζ_{32}	0.0207	0.0226	0.0213	0.0210	600
99	ζ_{32}	0.0175	0.0191	0.0184	0.0184	800
100	ζ_{32}	0.0161	0.0174	0.0165	0.0164	1000
101	ζ_{33}	0.0654	0.0721	0.0669	0.0677	200
102	ζ_{33}	0.0464	0.0508	0.0472	0.0474	400
103	ζ_{33}	0.0374	0.0414	0.0387	0.0381	600
104	ζ_{33}	0.0320	0.0355	0.0335	0.0331	800
105	ζ_{33}	0.0292	0.0318	0.0301	0.0297	1000

Table 13: Average standard errors calculated using the Empirical Fisher Information (EFI).

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
1	λ_{21}	0.0934	0.0987	0.0973	0.0946	200
2	λ_{21}	0.0644	0.0684	0.0664	0.0657	400
3	λ_{21}	0.0520	0.0557	0.0536	0.0531	600
4	λ_{21}	0.0449	0.0480	0.0463	0.0458	800
5	λ_{21}	0.0400	0.0428	0.0414	0.0409	1000
6	λ_{31}	0.1084	0.1147	0.1125	0.1096	200
7	λ_{31}	0.0745	0.0793	0.0771	0.0761	400
8	λ_{31}	0.0602	0.0645	0.0622	0.0615	600
9	λ_{31}	0.0520	0.0556	0.0536	0.0530	800
10	λ_{31}	0.0463	0.0496	0.0479	0.0474	1000
11	λ_{52}	0.0563	0.0595	0.0585	0.0572	200
12	λ_{52}	0.0389	0.0414	0.0401	0.0396	400
13	λ_{52}	0.0314	0.0336	0.0324	0.0321	600
14	λ_{52}	0.0271	0.0290	0.0280	0.0277	800
15	λ_{52}	0.0242	0.0259	0.0250	0.0247	1000
16	λ_{62}	0.0562	0.0593	0.0583	0.0571	200
17	λ_{62}	0.0387	0.0413	0.0400	0.0395	400
18	λ_{62}	0.0314	0.0335	0.0324	0.0320	600
19	λ_{62}	0.0271	0.0289	0.0279	0.0277	800
20	λ_{62}	0.0241	0.0258	0.0249	0.0247	1000
21	λ_{83}	0.4217	0.1648	0.1630	0.5464	200
22	λ_{83}	0.1070	0.1136	0.1102	0.1089	400
23	λ_{83}	0.0861	0.0919	0.0891	0.0882	600

(continue in the next page)

Table 13: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
24	λ_{83}	0.0745	0.0796	0.0767	0.0761	800
25	λ_{83}	0.0663	0.0709	0.0685	0.0678	1000
26	λ_{93}	0.3708	0.1490	0.1468	0.5254	200
27	λ_{93}	0.0966	0.1027	0.0994	0.0982	400
28	λ_{93}	0.0778	0.0829	0.0803	0.0795	600
29	λ_{93}	0.0672	0.0718	0.0692	0.0688	800
30	λ_{93}	0.0599	0.0640	0.0618	0.0611	1000
31	ψ_{11}	0.0694	0.0752	0.0712	0.0705	200
32	ψ_{11}	0.0478	0.0523	0.0494	0.0488	400
33	ψ_{11}	0.0386	0.0425	0.0401	0.0397	600
34	ψ_{11}	0.0333	0.0367	0.0346	0.0342	800
35	ψ_{11}	0.0297	0.0327	0.0309	0.0305	1000
36	ψ_{22}	0.0651	0.0754	0.0685	0.0669	200
37	ψ_{22}	0.0451	0.0527	0.0475	0.0467	400
38	ψ_{22}	0.0366	0.0429	0.0385	0.0379	600
39	ψ_{22}	0.0316	0.0371	0.0333	0.0327	800
40	ψ_{22}	0.0282	0.0332	0.0297	0.0292	1000
41	ψ_{33}	0.0632	0.0708	0.0658	0.0646	200
42	ψ_{33}	0.0437	0.0494	0.0457	0.0450	400
43	ψ_{33}	0.0355	0.0402	0.0370	0.0365	600
44	ψ_{33}	0.0306	0.0347	0.0320	0.0315	800
45	ψ_{33}	0.0273	0.0310	0.0286	0.0282	1000
46	ψ_{44}	0.0331	0.0369	0.0347	0.0339	200
47	ψ_{44}	0.0230	0.0258	0.0239	0.0236	400
48	ψ_{44}	0.0186	0.0210	0.0194	0.0192	600
49	ψ_{44}	0.0161	0.0181	0.0168	0.0165	800
50	ψ_{44}	0.0144	0.0162	0.0150	0.0148	1000
51	ψ_{55}	0.0330	0.0366	0.0344	0.0336	200
52	ψ_{55}	0.0229	0.0256	0.0238	0.0234	400
53	ψ_{55}	0.0185	0.0208	0.0193	0.0190	600
54	ψ_{55}	0.0160	0.0179	0.0167	0.0165	800
55	ψ_{55}	0.0143	0.0161	0.0149	0.0147	1000
56	ψ_{66}	0.0335	0.0374	0.0351	0.0343	200
57	ψ_{66}	0.0233	0.0262	0.0243	0.0239	400
58	ψ_{66}	0.0189	0.0214	0.0197	0.0194	600
59	ψ_{66}	0.0163	0.0184	0.0170	0.0168	800
60	ψ_{66}	0.0146	0.0165	0.0152	0.0150	1000
61	ψ_{77}	0.0613	0.0697	0.0642	0.0629	200
62	ψ_{77}	0.0425	0.0487	0.0445	0.0438	400

(continue in the next page)

Table 13: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
63	ψ_{77}	0.0345	0.0396	0.0361	0.0356	600
64	ψ_{77}	0.0298	0.0342	0.0312	0.0307	800
65	ψ_{77}	0.0266	0.0305	0.0279	0.0274	1000
66	ψ_{88}	0.0648	0.0710	0.0674	0.0661	200
67	ψ_{88}	0.0449	0.0495	0.0466	0.0461	400
68	ψ_{88}	0.0363	0.0402	0.0378	0.0373	600
69	ψ_{88}	0.0314	0.0348	0.0326	0.0322	800
70	ψ_{88}	0.0280	0.0310	0.0291	0.0288	1000
71	ψ_{99}	0.0619	0.0693	0.0645	0.0632	200
72	ψ_{99}	0.0428	0.0483	0.0447	0.0439	400
73	ψ_{99}	0.0347	0.0392	0.0362	0.0357	600
74	ψ_{99}	0.0300	0.0339	0.0313	0.0309	800
75	ψ_{99}	0.0268	0.0303	0.0280	0.0276	1000
76	ζ_{11}	0.0899	0.0969	0.0935	0.0914	200
77	ζ_{11}	0.0619	0.0674	0.0638	0.0631	400
78	ζ_{11}	0.0501	0.0547	0.0518	0.0514	600
79	ζ_{11}	0.0432	0.0473	0.0448	0.0443	800
80	ζ_{11}	0.0385	0.0423	0.0399	0.0395	1000
81	ζ_{21}	0.0486	0.0525	0.0512	0.0495	200
82	ζ_{21}	0.0336	0.0366	0.0347	0.0343	400
83	ζ_{21}	0.0272	0.0298	0.0282	0.0278	600
84	ζ_{21}	0.0234	0.0257	0.0243	0.0240	800
85	ζ_{21}	0.0209	0.0230	0.0217	0.0214	1000
86	ζ_{22}	0.0749	0.0837	0.0796	0.0767	200
87	ζ_{22}	0.0517	0.0582	0.0540	0.0531	400
88	ζ_{22}	0.0420	0.0474	0.0437	0.0431	600
89	ζ_{22}	0.0362	0.0409	0.0378	0.0372	800
90	ζ_{22}	0.0323	0.0365	0.0337	0.0332	1000
91	ζ_{31}	0.0396	0.0424	0.0414	0.0402	200
92	ζ_{31}	0.0273	0.0295	0.0282	0.0279	400
93	ζ_{31}	0.0222	0.0240	0.0229	0.0227	600
94	ζ_{31}	0.0190	0.0207	0.0198	0.0195	800
95	ζ_{31}	0.0170	0.0185	0.0176	0.0174	1000
96	ζ_{32}	0.0372	0.0400	0.0392	0.0379	200
97	ζ_{32}	0.0257	0.0278	0.0266	0.0263	400
98	ζ_{32}	0.0209	0.0226	0.0215	0.0213	600
99	ζ_{32}	0.0180	0.0195	0.0186	0.0184	800
100	ζ_{32}	0.0160	0.0174	0.0166	0.0164	1000
101	ζ_{33}	0.0677	0.0730	0.0707	0.0689	200

(continue in the next page)

Table 13: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
102	ζ_{33}	0.0468	0.0508	0.0485	0.0478	400
103	ζ_{33}	0.0379	0.0413	0.0393	0.0388	600
104	ζ_{33}	0.0327	0.0356	0.0339	0.0335	800
105	ζ_{33}	0.0292	0.0319	0.0303	0.0299	1000

Table 14: Average standard errors calculated using the Central Difference Method (CDM).

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
1	λ_{21}	0.0870	0.0935	0.0999	0.0887	200
2	λ_{21}	0.0611	0.0656	0.0633	0.0626	400
3	λ_{21}	0.0498	0.0536	0.0515	0.0510	600
4	λ_{21}	0.0431	0.0464	0.0445	0.0441	800
5	λ_{21}	0.0386	0.0414	0.0399	0.0395	1000
6	λ_{31}	0.0983	0.1056	0.1174	0.1002	200
7	λ_{31}	0.0690	0.0741	0.0715	0.0707	400
8	λ_{31}	0.0562	0.0605	0.0581	0.0575	600
9	λ_{31}	0.0487	0.0523	0.0502	0.0497	800
10	λ_{31}	0.0435	0.0467	0.0449	0.0445	1000
11	λ_{52}	0.0533	0.0770	0.0571	0.0546	200
12	λ_{52}	0.0376	0.0405	0.0389	0.0385	400
13	λ_{52}	0.0307	0.0330	0.0317	0.0314	600
14	λ_{52}	0.0266	0.0285	0.0274	0.0272	800
15	λ_{52}	0.0237	0.0255	0.0246	0.0243	1000
16	λ_{62}	0.0532	0.0755	0.0568	0.0544	200
17	λ_{62}	0.0375	0.0404	0.0388	0.0385	400
18	λ_{62}	0.0306	0.0329	0.0316	0.0314	600
19	λ_{62}	0.0265	0.0285	0.0274	0.0271	800
20	λ_{62}	0.0237	0.0255	0.0245	0.0243	1000
21	λ_{83}	0.1508	0.1626	0.1828	0.1543	200
22	λ_{83}	0.1055	0.1133	0.1090	0.1081	400
23	λ_{83}	0.0855	0.0919	0.0887	0.0878	600
24	λ_{83}	0.0742	0.0797	0.0766	0.0761	800
25	λ_{83}	0.0662	0.0711	0.0684	0.0678	1000
26	λ_{93}	0.1358	0.1466	0.1615	0.1388	200
27	λ_{93}	0.0951	0.1021	0.0981	0.0971	400
28	λ_{93}	0.0771	0.0827	0.0798	0.0790	600
29	λ_{93}	0.0668	0.0718	0.0690	0.0685	800
30	λ_{93}	0.0596	0.0640	0.0617	0.0610	1000
31	ψ_{11}	0.0657	0.0729	0.0759	0.0673	200
32	ψ_{11}	0.0459	0.0511	0.0477	0.0471	400

(continue in the next page)

Table 14: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
33	ψ_{11}	0.0373	0.0416	0.0389	0.0385	600
34	ψ_{11}	0.0323	0.0360	0.0337	0.0333	800
35	ψ_{11}	0.0288	0.0322	0.0301	0.0297	1000
36	ψ_{22}	0.0626	0.0733	0.0676	0.0649	200
37	ψ_{22}	0.0442	0.0518	0.0466	0.0459	400
38	ψ_{22}	0.0362	0.0423	0.0380	0.0375	600
39	ψ_{22}	0.0313	0.0366	0.0329	0.0325	800
40	ψ_{22}	0.0280	0.0328	0.0295	0.0290	1000
41	ψ_{33}	0.0602	0.0679	0.0669	0.0621	200
42	ψ_{33}	0.0425	0.0479	0.0444	0.0438	400
43	ψ_{33}	0.0346	0.0391	0.0362	0.0357	600
44	ψ_{33}	0.0300	0.0338	0.0313	0.0309	800
45	ψ_{33}	0.0268	0.0302	0.0280	0.0276	1000
46	ψ_{44}	0.0318	0.0360	0.0342	0.0328	200
47	ψ_{44}	0.0225	0.0255	0.0235	0.0232	400
48	ψ_{44}	0.0184	0.0208	0.0192	0.0190	600
49	ψ_{44}	0.0159	0.0180	0.0166	0.0164	800
50	ψ_{44}	0.0142	0.0161	0.0149	0.0147	1000
51	ψ_{55}	0.0317	0.0357	0.0338	0.0327	200
52	ψ_{55}	0.0224	0.0252	0.0234	0.0231	400
53	ψ_{55}	0.0183	0.0206	0.0191	0.0188	600
54	ψ_{55}	0.0158	0.0178	0.0165	0.0163	800
55	ψ_{55}	0.0142	0.0159	0.0148	0.0146	1000
56	ψ_{66}	0.0323	0.0365	0.0345	0.0333	200
57	ψ_{66}	0.0228	0.0258	0.0239	0.0236	400
58	ψ_{66}	0.0186	0.0211	0.0195	0.0192	600
59	ψ_{66}	0.0162	0.0183	0.0169	0.0167	800
60	ψ_{66}	0.0144	0.0163	0.0151	0.0149	1000
61	ψ_{77}	0.0591	0.0684	0.0653	0.0612	200
62	ψ_{77}	0.0418	0.0483	0.0439	0.0432	400
63	ψ_{77}	0.0341	0.0394	0.0358	0.0353	600
64	ψ_{77}	0.0295	0.0341	0.0310	0.0305	800
65	ψ_{77}	0.0264	0.0305	0.0277	0.0273	1000
66	ψ_{88}	0.0625	0.0695	0.0686	0.0644	200
67	ψ_{88}	0.0441	0.0490	0.0459	0.0454	400
68	ψ_{88}	0.0359	0.0399	0.0374	0.0370	600
69	ψ_{88}	0.0311	0.0345	0.0324	0.0320	800
70	ψ_{88}	0.0278	0.0308	0.0290	0.0286	1000
71	ψ_{99}	0.0597	0.0678	0.0645	0.0615	200

(continue in the next page)

Table 14: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
72	ψ_{99}	0.0421	0.0476	0.0440	0.0434	400
73	ψ_{99}	0.0343	0.0388	0.0359	0.0354	600
74	ψ_{99}	0.0297	0.0336	0.0311	0.0307	800
75	ψ_{99}	0.0266	0.0301	0.0278	0.0274	1000
76	ζ_{11}	0.0745	0.0795	0.0894	0.0758	200
77	ζ_{11}	0.0522	0.0560	0.0536	0.0532	400
78	ζ_{11}	0.0425	0.0456	0.0438	0.0435	600
79	ζ_{11}	0.0368	0.0395	0.0380	0.0377	800
80	ζ_{11}	0.0329	0.0353	0.0339	0.0336	1000
81	ζ_{21}	0.0394	0.0407	0.0411	0.0399	200
82	ζ_{21}	0.0280	0.0289	0.0283	0.0282	400
83	ζ_{21}	0.0229	0.0236	0.0231	0.0231	600
84	ζ_{21}	0.0198	0.0205	0.0201	0.0200	800
85	ζ_{21}	0.0177	0.0183	0.0179	0.0179	1000
86	ζ_{22}	0.0642	0.0715	0.0698	0.0660	200
87	ζ_{22}	0.0455	0.0505	0.0473	0.0467	400
88	ζ_{22}	0.0371	0.0413	0.0386	0.0381	600
89	ζ_{22}	0.0321	0.0358	0.0334	0.0330	800
90	ζ_{22}	0.0288	0.0320	0.0299	0.0296	1000
91	ζ_{31}	0.0352	0.0365	0.0383	0.0357	200
92	ζ_{31}	0.0249	0.0258	0.0252	0.0252	400
93	ζ_{31}	0.0204	0.0210	0.0206	0.0206	600
94	ζ_{31}	0.0176	0.0182	0.0179	0.0178	800
95	ζ_{31}	0.0158	0.0163	0.0160	0.0159	1000
96	ζ_{32}	0.0331	0.0344	0.0356	0.0336	200
97	ζ_{32}	0.0234	0.0243	0.0238	0.0237	400
98	ζ_{32}	0.0192	0.0199	0.0194	0.0194	600
99	ζ_{32}	0.0166	0.0172	0.0168	0.0168	800
100	ζ_{32}	0.0148	0.0154	0.0151	0.0150	1000
101	ζ_{33}	0.0623	0.0676	0.0768	0.0640	200
102	ζ_{33}	0.0440	0.0477	0.0455	0.0451	400
103	ζ_{33}	0.0360	0.0389	0.0371	0.0368	600
104	ζ_{33}	0.0311	0.0336	0.0322	0.0318	800
105	ζ_{33}	0.0278	0.0301	0.0288	0.0285	1000

Table 15: Probability coverage of confidence intervals based on the Empirical Fisher Information (EFI).

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
1	λ_{21}	0.9600	0.9546	0.9560	0.9562	200
2	λ_{21}	0.9564	0.9506	0.9532	0.9490	400

(continue in the next page)

Table 15: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
3	λ_{21}	0.9546	0.9412	0.9574	0.9514	600
4	λ_{21}	0.9496	0.9544	0.9546	0.9472	800
5	λ_{21}	0.9520	0.9478	0.9508	0.9512	1000
6	λ_{31}	0.9604	0.9498	0.9580	0.9572	200
7	λ_{31}	0.9542	0.9518	0.9524	0.9502	400
8	λ_{31}	0.9504	0.9480	0.9506	0.9514	600
9	λ_{31}	0.9548	0.9484	0.9544	0.9496	800
10	λ_{31}	0.9454	0.9514	0.9494	0.9538	1000
11	λ_{52}	0.9586	0.9560	0.9562	0.9538	200
12	λ_{52}	0.9524	0.9544	0.9578	0.9520	400
13	λ_{52}	0.9568	0.9500	0.9520	0.9540	600
14	λ_{52}	0.9544	0.9476	0.9506	0.9508	800
15	λ_{52}	0.9514	0.9474	0.9530	0.9472	1000
16	λ_{62}	0.9566	0.9576	0.9552	0.9572	200
17	λ_{62}	0.9572	0.9508	0.9548	0.9546	400
18	λ_{62}	0.9558	0.9504	0.9464	0.9500	600
19	λ_{62}	0.9532	0.9486	0.9474	0.9502	800
20	λ_{62}	0.9510	0.9484	0.9536	0.9450	1000
21	λ_{83}	0.9566	0.9504	0.9564	0.9534	200
22	λ_{83}	0.9552	0.9510	0.9562	0.9568	400
23	λ_{83}	0.9522	0.9466	0.9536	0.9536	600
24	λ_{83}	0.9542	0.9542	0.9510	0.9550	800
25	λ_{83}	0.9524	0.9512	0.9550	0.9482	1000
26	λ_{93}	0.9566	0.9556	0.9566	0.9564	200
27	λ_{93}	0.9542	0.9538	0.9528	0.9514	400
28	λ_{93}	0.9522	0.9554	0.9548	0.9488	600
29	λ_{93}	0.9536	0.9510	0.9520	0.9514	800
30	λ_{93}	0.9520	0.9544	0.9504	0.9532	1000
31	ψ_{11}	0.9532	0.9530	0.9552	0.9616	200
32	ψ_{11}	0.9578	0.9508	0.9558	0.9544	400
33	ψ_{11}	0.9484	0.9494	0.9534	0.9530	600
34	ψ_{11}	0.9500	0.9502	0.9560	0.9474	800
35	ψ_{11}	0.9554	0.9522	0.9502	0.9492	1000
36	ψ_{22}	0.9506	0.9476	0.9451	0.9544	200
37	ψ_{22}	0.9538	0.9506	0.9512	0.9514	400
38	ψ_{22}	0.9474	0.9518	0.9518	0.9450	600
39	ψ_{22}	0.9502	0.9508	0.9482	0.9488	800
40	ψ_{22}	0.9498	0.9506	0.9440	0.9466	1000
41	ψ_{33}	0.9534	0.9464	0.9502	0.9496	200

(continue in the next page)

Table 15: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
42	ψ_{33}	0.9506	0.9506	0.9506	0.9494	400
43	ψ_{33}	0.9524	0.9530	0.9528	0.9476	600
44	ψ_{33}	0.9458	0.9558	0.9512	0.9500	800
45	ψ_{33}	0.9512	0.9518	0.9474	0.9532	1000
46	ψ_{44}	0.9526	0.9560	0.9498	0.9566	200
47	ψ_{44}	0.9472	0.9510	0.9540	0.9502	400
48	ψ_{44}	0.9544	0.9514	0.9486	0.9558	600
49	ψ_{44}	0.9532	0.9504	0.9510	0.9482	800
50	ψ_{44}	0.9520	0.9502	0.9528	0.9486	1000
51	ψ_{55}	0.9550	0.9492	0.9480	0.9544	200
52	ψ_{55}	0.9542	0.9500	0.9542	0.9522	400
53	ψ_{55}	0.9504	0.9484	0.9472	0.9516	600
54	ψ_{55}	0.9560	0.9510	0.9524	0.9478	800
55	ψ_{55}	0.9550	0.9504	0.9554	0.9500	1000
56	ψ_{66}	0.9506	0.9538	0.9514	0.9524	200
57	ψ_{66}	0.9536	0.9494	0.9488	0.9508	400
58	ψ_{66}	0.9512	0.9516	0.9548	0.9488	600
59	ψ_{66}	0.9458	0.9506	0.9472	0.9492	800
60	ψ_{66}	0.9502	0.9534	0.9546	0.9478	1000
61	ψ_{77}	0.9536	0.9474	0.9441	0.9540	200
62	ψ_{77}	0.9480	0.9480	0.9568	0.9454	400
63	ψ_{77}	0.9488	0.9530	0.9468	0.9436	600
64	ψ_{77}	0.9516	0.9442	0.9506	0.9504	800
65	ψ_{77}	0.9520	0.9508	0.9510	0.9488	1000
66	ψ_{88}	0.9598	0.9586	0.9580	0.9562	200
67	ψ_{88}	0.9538	0.9500	0.9546	0.9562	400
68	ψ_{88}	0.9540	0.9494	0.9542	0.9518	600
69	ψ_{88}	0.9562	0.9560	0.9530	0.9504	800
70	ψ_{88}	0.9526	0.9504	0.9530	0.9498	1000
71	ψ_{99}	0.9564	0.9554	0.9514	0.9528	200
72	ψ_{99}	0.9514	0.9518	0.9550	0.9530	400
73	ψ_{99}	0.9474	0.9550	0.9528	0.9556	600
74	ψ_{99}	0.9510	0.9482	0.9450	0.9490	800
75	ψ_{99}	0.9526	0.9498	0.9528	0.9462	1000
76	ζ_{11}	0.9502	0.9446	0.9500	0.9558	200
77	ζ_{11}	0.9554	0.9442	0.9504	0.9494	400
78	ζ_{11}	0.9480	0.9502	0.9466	0.9494	600
79	ζ_{11}	0.9548	0.9456	0.9542	0.9452	800
80	ζ_{11}	0.9436	0.9548	0.9484	0.9504	1000

(continue in the next page)

Table 15: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
81	ζ_{21}	0.9562	0.9494	0.9546	0.9540	200
82	ζ_{21}	0.9516	0.9492	0.9510	0.9524	400
83	ζ_{21}	0.9522	0.9470	0.9528	0.9526	600
84	ζ_{21}	0.9540	0.9494	0.9504	0.9524	800
85	ζ_{21}	0.9500	0.9496	0.9484	0.9518	1000
86	ζ_{22}	0.9596	0.9556	0.9506	0.9520	200
87	ζ_{22}	0.9602	0.9506	0.9508	0.9508	400
88	ζ_{22}	0.9530	0.9520	0.9482	0.9514	600
89	ζ_{22}	0.9512	0.9520	0.9514	0.9502	800
90	ζ_{22}	0.9524	0.9476	0.9458	0.9452	1000
91	ζ_{31}	0.9534	0.9472	0.9538	0.9466	200
92	ζ_{31}	0.9504	0.9512	0.9458	0.9518	400
93	ζ_{31}	0.9536	0.9490	0.9452	0.9474	600
94	ζ_{31}	0.9498	0.9516	0.9538	0.9472	800
95	ζ_{31}	0.9436	0.9520	0.9520	0.9436	1000
96	ζ_{32}	0.9512	0.9534	0.9562	0.9532	200
97	ζ_{32}	0.9514	0.9516	0.9504	0.9518	400
98	ζ_{32}	0.9518	0.9500	0.9528	0.9534	600
99	ζ_{32}	0.9554	0.9510	0.9528	0.9516	800
100	ζ_{32}	0.9476	0.9502	0.9490	0.9530	1000
101	ζ_{33}	0.9554	0.9498	0.9532	0.9494	200
102	ζ_{33}	0.9480	0.9496	0.9562	0.9494	400
103	ζ_{33}	0.9494	0.9480	0.9506	0.9526	600
104	ζ_{33}	0.9518	0.9526	0.9482	0.9476	800
105	ζ_{33}	0.9478	0.9458	0.9516	0.9524	1000

Table 16: Probability coverage of confidence intervals based on the Central Difference Method (CDM).

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
1	λ_{21}	0.9482	0.9426	0.9523	0.9435	200
2	λ_{21}	0.9450	0.9392	0.9450	0.9388	400
3	λ_{21}	0.9456	0.9338	0.9488	0.9442	600
4	λ_{21}	0.9394	0.9472	0.9444	0.9378	800
5	λ_{21}	0.9412	0.9406	0.9392	0.9406	1000
6	λ_{31}	0.9390	0.9318	0.9500	0.9371	200
7	λ_{31}	0.9348	0.9334	0.9328	0.9328	400
8	λ_{31}	0.9338	0.9326	0.9352	0.9354	600
9	λ_{31}	0.9418	0.9340	0.9364	0.9336	800
10	λ_{31}	0.9314	0.9370	0.9330	0.9394	1000
11	λ_{52}	0.9480	0.9490	0.9530	0.9445	200

(continue in the next page)

Table 16: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
12	λ_{52}	0.9458	0.9482	0.9524	0.9468	400
13	λ_{52}	0.9508	0.9470	0.9450	0.9508	600
14	λ_{52}	0.9472	0.9450	0.9464	0.9450	800
15	λ_{52}	0.9482	0.9436	0.9496	0.9432	1000
16	λ_{62}	0.9462	0.9492	0.9516	0.9479	200
17	λ_{62}	0.9506	0.9446	0.9492	0.9506	400
18	λ_{62}	0.9502	0.9468	0.9418	0.9454	600
19	λ_{62}	0.9486	0.9440	0.9418	0.9466	800
20	λ_{62}	0.9478	0.9462	0.9506	0.9412	1000
21	λ_{83}	0.9498	0.9482	0.9455	0.9513	200
22	λ_{83}	0.9516	0.9490	0.9548	0.9552	400
23	λ_{83}	0.9504	0.9472	0.9548	0.9540	600
24	λ_{83}	0.9532	0.9548	0.9510	0.9534	800
25	λ_{83}	0.9518	0.9514	0.9550	0.9490	1000
26	λ_{93}	0.9488	0.9504	0.9457	0.9475	200
27	λ_{93}	0.9522	0.9530	0.9500	0.9480	400
28	λ_{93}	0.9520	0.9548	0.9540	0.9472	600
29	λ_{93}	0.9522	0.9496	0.9528	0.9524	800
30	λ_{93}	0.9518	0.9546	0.9506	0.9526	1000
31	ψ_{11}	0.9450	0.9480	0.9732	0.9535	200
32	ψ_{11}	0.9490	0.9492	0.9480	0.9484	400
33	ψ_{11}	0.9402	0.9474	0.9486	0.9476	600
34	ψ_{11}	0.9450	0.9466	0.9476	0.9410	800
35	ψ_{11}	0.9490	0.9488	0.9442	0.9410	1000
36	ψ_{22}	0.9430	0.9418	0.9443	0.9499	200
37	ψ_{22}	0.9466	0.9474	0.9476	0.9468	400
38	ψ_{22}	0.9442	0.9494	0.9508	0.9422	600
39	ψ_{22}	0.9484	0.9492	0.9468	0.9466	800
40	ψ_{22}	0.9482	0.9488	0.9422	0.9452	1000
41	ψ_{33}	0.9470	0.9358	0.9505	0.9441	200
42	ψ_{33}	0.9438	0.9426	0.9448	0.9446	400
43	ψ_{33}	0.9452	0.9440	0.9474	0.9428	600
44	ψ_{33}	0.9418	0.9490	0.9464	0.9468	800
45	ψ_{33}	0.9480	0.9470	0.9446	0.9500	1000
46	ψ_{44}	0.9450	0.9494	0.9516	0.9511	200
47	ψ_{44}	0.9436	0.9502	0.9498	0.9490	400
48	ψ_{44}	0.9520	0.9490	0.9482	0.9552	600
49	ψ_{44}	0.9510	0.9486	0.9484	0.9448	800
50	ψ_{44}	0.9500	0.9490	0.9510	0.9484	1000

(continue in the next page)

Table 16: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
51	ψ_{55}	0.9482	0.9442	0.9473	0.9495	200
52	ψ_{55}	0.9502	0.9460	0.9522	0.9490	400
53	ψ_{55}	0.9494	0.9454	0.9450	0.9502	600
54	ψ_{55}	0.9544	0.9484	0.9506	0.9466	800
55	ψ_{55}	0.9548	0.9504	0.9528	0.9484	1000
56	ψ_{66}	0.9442	0.9480	0.9514	0.9473	200
57	ψ_{66}	0.9496	0.9482	0.9460	0.9474	400
58	ψ_{66}	0.9504	0.9492	0.9520	0.9474	600
59	ψ_{66}	0.9438	0.9470	0.9464	0.9484	800
60	ψ_{66}	0.9480	0.9520	0.9528	0.9466	1000
61	ψ_{77}	0.9474	0.9444	0.9523	0.9491	200
62	ψ_{77}	0.9446	0.9464	0.9548	0.9452	400
63	ψ_{77}	0.9450	0.9512	0.9478	0.9422	600
64	ψ_{77}	0.9490	0.9454	0.9496	0.9482	800
65	ψ_{77}	0.9512	0.9500	0.9496	0.9484	1000
66	ψ_{88}	0.9552	0.9556	0.9700	0.9509	200
67	ψ_{88}	0.9510	0.9486	0.9542	0.9534	400
68	ψ_{88}	0.9530	0.9480	0.9544	0.9510	600
69	ψ_{88}	0.9542	0.9538	0.9498	0.9494	800
70	ψ_{88}	0.9516	0.9488	0.9518	0.9494	1000
71	ψ_{99}	0.9500	0.9516	0.9597	0.9473	200
72	ψ_{99}	0.9474	0.9488	0.9528	0.9500	400
73	ψ_{99}	0.9442	0.9526	0.9522	0.9544	600
74	ψ_{99}	0.9506	0.9472	0.9436	0.9472	800
75	ψ_{99}	0.9508	0.9482	0.9518	0.9460	1000
76	ζ_{11}	0.9103	0.8866	0.9019	0.9046	200
77	ζ_{11}	0.9040	0.8918	0.9004	0.9042	400
78	ζ_{11}	0.8986	0.9006	0.9026	0.9056	600
79	ζ_{11}	0.9102	0.8992	0.9088	0.8982	800
80	ζ_{11}	0.9026	0.8986	0.8992	0.9048	1000
81	ζ_{21}	0.9039	0.8706	0.9008	0.8952	200
82	ζ_{21}	0.9006	0.8778	0.8912	0.9046	400
83	ζ_{21}	0.9090	0.8762	0.8984	0.9016	600
84	ζ_{21}	0.9104	0.8862	0.8996	0.9004	800
85	ζ_{21}	0.9004	0.8780	0.9002	0.8964	1000
86	ζ_{22}	0.9253	0.9152	0.9212	0.9185	200
87	ζ_{22}	0.9234	0.9116	0.9148	0.9204	400
88	ζ_{22}	0.9240	0.9168	0.9158	0.9240	600
89	ζ_{22}	0.9202	0.9188	0.9144	0.9196	800

(continue in the next page)

Table 16: continued from the previous page

	Parameter	MCFA-N	MCFA-t	MCFA-CN	MCFA-SL	Sample size
90	ζ_{22}	0.9242	0.9128	0.9200	0.9138	1000
91	ζ_{31}	0.9280	0.9056	0.9390	0.9177	200
92	ζ_{31}	0.9300	0.9174	0.9160	0.9254	400
93	ζ_{31}	0.9342	0.9176	0.9220	0.9240	600
94	ζ_{31}	0.9312	0.9184	0.9326	0.9188	800
95	ζ_{31}	0.9242	0.9208	0.9278	0.9190	1000
96	ζ_{32}	0.9280	0.9132	0.9267	0.9259	200
97	ζ_{32}	0.9310	0.9192	0.9210	0.9336	400
98	ζ_{32}	0.9328	0.9138	0.9300	0.9278	600
99	ζ_{32}	0.9350	0.9198	0.9246	0.9284	800
100	ζ_{32}	0.9294	0.9108	0.9260	0.9286	1000
101	ζ_{33}	0.9364	0.9340	0.9179	0.9353	200
102	ζ_{33}	0.9338	0.9320	0.9430	0.9350	400
103	ζ_{33}	0.9370	0.9354	0.9380	0.9416	600
104	ζ_{33}	0.9412	0.9400	0.9384	0.9372	800
105	ζ_{33}	0.9390	0.9332	0.9424	0.9392	1000