A Monte Carlo Study of Structural Equation Models for Finite Mixtures [*]

John Williams [†]

Dirk Temme [‡]

Lutz Hildebrandt

June 3, 2002

Abstract

Empirical applications of structural equation modeling (SEM) typically rest on the assumption that the analysed sample is homogenous with respect to the underlying structural model or that homogenous subsamples have been formed based on a priori knowledge. However, researchers often are ignorant about the true causes of heterogeneity and thus risk to produce misleading results. Using a sequential procedure of cluster analysis in combination with multi-group SEM has been shown to be inappropriate to solve the problem of unobserved heterogeneity. Recently, two encouraging approaches have been developed in this regard: (1) Finite mixtures of structural equation models and (2) hierarchical Bayesian estimation. In this paper, we focus exclusively on the MECOSA approach to finite normal mixtures subject to conditional mean and covariance structures. Since not much is known about the performance of MECOSA, which is both a specific model and a software, we present the results of an extensive Monte Carlo simulation. It was found that MECOSA performed best where homogenous groups were present in the data in equal proportions and in conjunction with rather large differences in parameters across the groups. MECOSA performed worse when the proportions were unequal and parameters were relatively close together across groups. Of the three estimation methods available in MECOSA the two-stage minimum distance estimation (MDE) in general performed worse than the alternative EM algorithms (EM and EMG). This effect was especially pronounced under conditions of close parameters and unequal group proportions. Above that, for these conditions the modified likelihood ratio test turned out to be inappropriate in the three groups case.

1 Introduction

Structural equation modeling (SEM) or mean and covariance structure analysis (both terms will be used interchangeably) is an established, widely used methodology in the social sciences (e.g., sociology, psychology, marketing). Its key characteristic is the simultaneous estimation of relationships among unobserved or latent variables, which represent the hypothetical constructs (e.g., attitudes, values, customer satisfaction) of a theory, and between the latent variables and their observed indicators (see for example Bollen, 1989). Although the individual manifest variables are conceptualised as imperfect measures of the hypothetical constructs measurement errors can be controlled by using multiple indicators (e.g., Bagozzi, Yi & Nassen, 1999).

Empirical applications of SEM typically rest on the assumption that either the analysed sample is homogenous with respect to the underlying structural model which has generated the data or that sufficiently homogenous subsamples have been formed based on theoretical considerations about possible sources of heterogeneity. For example, a researcher might suspect that dependent on their previous buying behaviour (e.g., buying intensity) subjects in a survey on customers' attitudes toward specific brands show systematic differences in their responses. This can have an effect on both the measurement part (e.g., latent variable means differ across the subjects) and the structural part (e.g., relationships between the latent variables differ across the subjects) of a structural equation model (see for example Jagpal, 1999). In this case, the sample might be divided into segments of light and heavy users. Multigroup analysis, an option available in most SEM software packages (e.g., AMOS, EQS, LISREL, Mplus), may then be used to test for actual differences between the segments and to estimate (partially) group-specific models if necessary. As useful as such an analysis can be, often observable characteristics (e.g., age, gender, loyalty) are insufficient to identify homogenous groups or researchers are simply ignorant of the true sources of heterogeneity. If unobserved heterogeneity in an empirical study is substantial, parameter estimates can be seriously distorted, thus leading to erroneous inferences (Jedidi, Jagpal & DeSarbo, 1997a).

As an ad hoc solution to this problem a sequential procedure might be used. In the first step "homogenous" subgroups are formed by applying cluster analysis (e.g., k-means clustering) on the data, followed by a multi-group SEM analysis of the identified clusters in the second step. However, both theoretical considerations as well as simulation evidence lead to the conclusion that this approach suffers from serious shortcomings. Whereas researchers typically develop specific hypotheses about the relationships between the variables of interest, which is mirrored in the structural equation model tested in the second step, traditional cluster analysis assumes independence among these variables. In principal, data-reduction techniques (e.g., principal component analysis) can precede the cluster analysis in the case of highly correlated variables (e.g., indicators of the same latent variable), but since the former is based on the assumption of homogeneous data, this approach is conceptually flawed. In addition, simulation studies have shown that parameter recovery in the two-step procedure is rather weak (Görz, Hildebrandt & Annacker, 2000, Jedidi, Jagpal & DeSarbo, 1997a).

Meanwhile two encouraging approaches for modeling unobserved heterogeneity in SEM have

been developed. The finite mixture approach to SEM (e.g., Jedidi, Jagpal & DeSarbo, 1997a, Arminger, Stein & Wittenberg, 1999) assumes that a heterogenous population consists of a finite number of groups or components each characterized by a specific multivariate distribution. In contrast to multi-group analysis the number of components and the component-specific distributions are in general unknown and have to be estimated. Whereas heterogeneity in finite mixture analysis is represented by a discrete distribution, the hierarchical Bayesian approach (Ansari, Jedidi & Jagpal, 2000) rests upon the idea that individual-level parameters follow a continuous heterogeneity distribution. Given that multiple observations are available for each subject, this methodology allows to estimate parameters (e.g., latent variable means) at the level of individual subjects. From a methodological viewpoint both approaches have their pros and cons (for a general discussion see for example Wedel, Kamakura & Böckenholt, 2000, Allenby & Rossi, 1999). For example, if unobserved heterogeneity is such that subjects do not form relatively homogenous subgroups finite mixture analysis might produce misleading results. On the other side, assuming that individual-level parameters follow a specific distribution (e.g., a normal distribution) can be problematic if the true distribution is, for example, multi-modal. In addition to this, unlike finite mixture SEM the hierarchical Bayesian approach requires that multiple observations for the subjects are available. From a managerial viewpoint, finite mixture analysis seems especially suitable in the context of segmenting large consumer goods markets whereas hierarchical Bayesian estimation seems more appropriate in the case of direct one-to-one marketing.

In this paper, we refrain from comparing both methods but exclusively focus on the validity of the finite mixture approach to SEM when unobserved population heterogeneity is indeed correctly captured by a discrete distribution. Compared to the vast amount of simulation studies which exists for "regular" structural equation models relatively few is known about the performance of finite mixture SEM. For this reason, we conducted an extensive Monte Carlo simulation study into the properties of MECOSA (MEan and COvariance Structure Analysis, Arminger, Wittenberg & Schepers, 1996). MECOSA is both a specific model and a multi-purpose software which among other things facilitates the estimation of conditional and unconditional finite mixture models.

The paper is organised as follows: First, we give a brief introduction to finite mixture structural equation modeling and explain the main differences between the proposed models. Second, the MECOSA approach to conditional finite mixtures is presented in more depth, including the model and the estimation procedures implemented in MECOSA. Third, the experimental design and the results of the simulation study are presented. Finally, we summarise the results of our study and discuss their implications for empirical applications of MECOSA. We also suggest directions for further research.

2   Finite Mixture Structural Equation Models

2.1   Conditional Versus Unconditional Models for Normal Mixtures

The finite mixture approach assumes that a heterogeneous population consists of a limited, but generally unknown number of subpopulations or groups each characterised by a specific distribution (for general descriptions of this methodology see for example Everitt & Hand, 1981, McLachlan & Basford, 1988, McLachlan & Peel, 2000). The purpose of finite mixture analysis therefore is to determine the number of groups in the population and to estimate the component-specific parameters. Based on these estimates probabilities of group membership can be calculated and subjects be assigned to the different groups. This method has found increasing interest over the last decades especially since the seminal paper by Dempster, Laird & Rubin (1977) on the EM algorithm. But though Blåfield (1980) proposed a confirmatory factor model for finite normal mixtures even more than ten years ago, it was only recently that the application of finite mixtures to structural equation models has gained momentum.

With respect to the distributional assumptions for the observed variables, two types of finite mixture SEM can be distinguished. Unconditional models (Yung, 1994, 1997, Jedidi, Jagpal & DeSarbo, 1997a, Dolan & van der Maas, 1998) presume that the endogenous and exogenous variables follow a multivariate normal distribution within the different mixture components. In contrast, for conditional models (Arminger & Stein, 1997, Arminger, Stein & Wittenberg, 1999, Muthén & Shedden, 1999) the somewhat weaker assumption applies that the dependent variables are normally distributed given some arbitrarily distributed exogenous regressor variables (e.g., socio-demographics like age, gender, or occupation). Whereas in the single or multi-group case the estimators (e.g., ML or GLS) remain consistent even if the assumption of multivariate normality is violated because of non-normal regressors (see for example Bollen, 1989), obviously this does not hold for finite mixture structural equation models as simulation studies have shown (Arminger, Stein & Wittenberg, 1999). Because in empirical applications multivariate normality is more the exception than the rule, conditional models seem to have more practical relevance.

To the best of our knowledge only two general software programmes for the estimation of conditional finite mixture SEM are presently available (MECOSA, Arminger, Wittenberg & Schepers, 1996, and Mplus, Muthén & Muthén, 1998). We decided to employ MECOSA in our simulation study largely because it offers three estimation methods, one of which (the EM algorithm to be described later on) is very similar to the algorithm used in Mplus. Thus, we expect that our simulation results to a large extent also apply to the Mplus software.

2.2   The MECOSA Approach to Conditional Finite Mixture SEM

2.2.1   MECOSA Model for Conditional Finite Normal Mixtures

The following description of the MECOSA model for conditional normal mixtures is drawn closely from (Arminger, Stein & Wittenberg, 1999). For subject i, $i = 1, 2, \ldots, n$, let $\mathbf{y}_i$ be a p-dimensional vector of continuous dependent random variables and $\mathbf{x}_i$ a q-dimensional vector

of independent variables. The latter may include continuous and/or dummy-type variables. The sample points $\mathbf{z}_i = (\mathbf{y}_i, \mathbf{x}_i)$ are i. i. d. with $h(\mathbf{y}_i, \mathbf{x}_i) = f(\mathbf{y}_i | \mathbf{x}_i) \cdot g(\mathbf{x}_i)$, where $f(\mathbf{y}_i | \mathbf{x}_i)$ is the conditional density of $\mathbf{y}_i$ given $\mathbf{x}_i$ and $g(\mathbf{x}_i)$ is the marginal density of the exogenous variables. If the dependent variables $\mathbf{y}_i$ are multivariate normal in each component conditional on the regressors $\mathbf{x}_i$, the conditional density is given by the following mixture:

$$f(\mathbf{y}_i | \mathbf{x}_i) = \pi_1 \phi(\mathbf{y}_i; \boldsymbol{\mu}_{i1}; \boldsymbol{\Sigma}_1) + \pi_2 \phi(\mathbf{y}_i; \boldsymbol{\mu}_{i2}; \boldsymbol{\Sigma}_2) + \cdots + \pi_K \phi(\mathbf{y}_i; \boldsymbol{\mu}_{iK}; \boldsymbol{\Sigma}_K), \tag{1}$$

where $\pi_k, k = 1, 2, \ldots, K$, are the mixing proportions for the K components of the mixture and ϕ denotes the multivariate normal density with mean vector $\boldsymbol{\mu}_{ik}$ and covariance matrix $\boldsymbol{\Sigma}_k$. The mixing proportions are subject to the following constraints: $\pi_k \geq 0$ and $\sum_{k=1}^{K} \pi_k = 1$. The conditional mean $E(\mathbf{y}_i | \mathbf{x}_i, k)$ is specified as a reduced form multivariate linear model

$$\boldsymbol{\mu}_{ik} = \boldsymbol{\gamma}_k + \boldsymbol{\Pi}_k \mathbf{x}_i, \tag{2}$$

where $\boldsymbol{\gamma}_k$ is a p-dimensional vector of regression constants, $\boldsymbol{\Pi}_k$ is a $p \times q$ matrix of regression parameters, and the conditional covariance matrix $\boldsymbol{\Sigma}_k$ contains the variances and covariances of the regression residuals. The elements of $\boldsymbol{\gamma}_k$, $\boldsymbol{\Pi}_k$ and $\boldsymbol{\Sigma}_k$ are called *reduced form parameters* and are collected in an r-dimensional vector $\boldsymbol{\kappa}$. Augmenting the vector $\boldsymbol{\kappa}$ by the non-redundant mixing proportions yields the vector $\boldsymbol{\kappa}^* = ((\pi_1, \ldots, \pi_{k-1}), \boldsymbol{\kappa}')')$.

The conditional means and conditional covariance matrices are parameterised by component-specific mean and covariance structure models, where the free, *fundamental parameters* are collected in the d-dimensional vector $\boldsymbol{\vartheta}$. The conditional density can thus be written as

$$f(\mathbf{y}_i | \mathbf{x}_i) = \pi_1 \phi(\mathbf{y}_i; \boldsymbol{\mu}_{i1}(\boldsymbol{\vartheta}); \boldsymbol{\Sigma}_1(\boldsymbol{\vartheta})) + \cdots + \pi_K \phi(\mathbf{y}_i; \boldsymbol{\mu}_{iK}(\boldsymbol{\vartheta}); \boldsymbol{\Sigma}_K(\boldsymbol{\vartheta})). \tag{3}$$

Within each component the conditional mean and the conditional covariance matrix can be parameterised as, for example, a conditional LISREL model. In this case, the structural equation model for the latent variables is

$$\boldsymbol{\eta}_i | (\mathbf{x}_i, k) = \mathbf{B}_k \boldsymbol{\eta}_i + \boldsymbol{\Gamma}_k \mathbf{x}_i + \boldsymbol{\zeta}_i^{(k)}, \tag{4}$$

with $\boldsymbol{\zeta}_i^{(k)} \sim N(\mathbf{0}, \boldsymbol{\Psi}_k)$. The latent endogenous variables are connected to the observed endogenous indicator variables by group-specific factor analytic models

$$\mathbf{y}_i = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \boldsymbol{\eta}_i^{(k)} + \boldsymbol{\epsilon}_i^{(k)}, \tag{5}$$

where $\boldsymbol{\eta}_i^{(k)} = \boldsymbol{\eta}_i | (\mathbf{x}_i, k)$ and $\boldsymbol{\epsilon}_i^{(k)} \sim N(\mathbf{0}, \boldsymbol{\Theta}_k)$. For the conditional mean of the endogenous variables this implies

$$E(\mathbf{y}_i | \mathbf{x}_i, k) = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k (\mathbf{I} - \mathbf{B}_k)^{-1} \boldsymbol{\Gamma}_k \mathbf{x}_i = \boldsymbol{\gamma}_k + \boldsymbol{\Pi}_k \mathbf{x}_i, \tag{6}$$

where $\boldsymbol{\gamma}_k = \boldsymbol{\nu}_k$ and $\boldsymbol{\Pi}_k = \boldsymbol{\Lambda}_k (\mathbf{I} - \mathbf{B}_k)^{-1} \boldsymbol{\Gamma}_k$. The conditional covariance matrix is specified as

$$V(\mathbf{y}_i | \mathbf{x}_i, k) = \boldsymbol{\Lambda}_k (\mathbf{I} - \boldsymbol{B}_k)^{-1} \boldsymbol{\Psi}_k (\mathbf{I} - \boldsymbol{B}_k)^{-1'} \boldsymbol{\Lambda}_k' + \boldsymbol{\Theta}_k = \boldsymbol{\Sigma}_k. \tag{7}$$

Given that the individual component-specific mean and covariance structure models are identified (this is analogous to the multi-group case), normality of the dependent variables in each

component given the regressors is sufficient and necessary to establish the identifiability of finite mixture structural equation models (Jedidi, Jagpal & DeSarbo, 1997a). Here, we implicitly assume that the restriction $\pi_1 \leq \pi_2 \leq \ldots \leq \pi_K$ (Aitkin & Rubin, 1985) is imposed to solve the problem of arbitrary permutations of the components labels (McLachlan & Basford, 1988).

2.2.2 Parameter Estimation

MECOSA currently offers three estimation methods (for a detailed description see Arminger, Stein & Wittenberg, 1999): Minimum Distance Estimation (MDE), Direct EM (EM) and EM Gradient (EMG).

The MDE proceeds in two stages. In the first stage, for a given number of components an expectation-minimization (EM) algorithm is used to estimate the reduced form parameters γ_k, Π_k and Σ_k of an unrestricted regression of $\mathbf{y}$ on $\mathbf{x}$. Following, the asymptotic covariance matrix Ω^* of the vector $\hat{\kappa}^*$ is determined, which includes the asymptotic covariance matrix Ω of the estimated reduced form parameters in $\hat{\kappa}$ as a submatrix. In the second stage, minimum distance estimation is used to determine the group specific-fundamental parameters $\hat{\vartheta}$ based on the reduced form parameter estimates and their asymptotic covariance matrix. This is done by minimizing the function

$$Q(\vartheta) = [\hat{\kappa} - \kappa(\vartheta)]' \, \hat{\Omega}^{-1} [\hat{\kappa} - \kappa(\vartheta)] \,. \tag{8}$$

Under the null hypothesis that the component-specific mean and covariance structure models represent the true underlying models which gave rise to the reduced form parameters, i. e. $\kappa = \kappa(\vartheta)$, we obtain a test statistic $Q(\hat{\vartheta})$ which is asymptotically distributed as a χ^2 variable with $r - d$ degrees of freedom.

In contrast to MDE, maximisation of the log-likelihood function during the M step of the EM and the EMG algorithms is directly performed with respect to the fundamental parameters ϑ instead of the reduced form parameters. EM and EMG only differ with respect to the number of iterations in the M step: In EMG only one iteration is performed in each M step, whereas in the EM algorithm iterations proceed until convergence is achieved. Starting values for EM and EMG can be provided directly or can be generated by using the parameter estimates of the MDE procedure.

Since in the M-step of MDE weighted regression is used instead of an iterative procedure like in EM and EMG the former is much faster than both EM and EMG, with EM beeing the slowest. It is not the purpose of this paper to explain or debate pros and cons of each method (see Arminger, Stein & Wittenberg, 1999), but simply to examine their properties in a simulation. However, it should be noted that Arminger, Stein & Wittenberg (1999) recommend that MDE should be used when the model has few parameters, and EM Gradient in combination with the first stage of the MDE procedure when the structural equation model is large.

2.2.3 Deciding on the Number of Groups

Because in applications of the finite mixture approach the number of components is typically unknown, statistical criteria are required to decide on this issue. Violations of the regularity conditions prevent the application of a conventional likelihood ratio test in this case (see McLachlan & Basford, 1988). Modifications of this test (Wolfe, 1971), a parametric bootstrap of the likelihhod ratio (Aitkin, Anderson & Hinde, 1981, McLachlan, 1987) and information criteria like AIC, BIC or CAIC (see for example Jedidi, Jagpal & DeSarbo, 1997b) have been proposed instead. In MECOSA both an extension of the ad hoc test of Wolfe (1971) to more than two groups (as default) and the simplified parametric bootstrap procedure proposed by Aitkin, Anderson & Hinde (1981) are implemented (Arminger, Stein & Wittenberg, 1999).

The ad hoc procedure tests the null hypothesis (H_0) that the sample consists of K components against the alternative hypothesis (H_1) of $K + 1$ components. The null hypothesis is rejected at a given significance level α if

$$\lambda \ > \ \chi^2_{(K+1)(r_1-r_0)}, \tag{9}$$

where λ is the conventional likelihood ratio statistic, r_1 is the number of parameters under H_1 and r_0 is the number of parameters under H_1, both without the mixing proportions.

Although there is some evidence (e.g., Everitt, 1981) that the ad hoc test does not perform well and that bootstrapping seems to be a more promising way forward (McLachlan, 1987, Feng & McCulloch, 1996), the ad hoc test has been used in this study because of the hitherto prohibitive cost in terms of computing requirements even for only moderately complex models. With increasing computing power this may be less of a problem in future simulation studies.

3 Simulation Study

The purpose of our Monte Carlo simulation is to investigate if MECOSA is an appropriate tool to control for unobserved heterogeneity in SEM when the population consists of a finite but unknown number of relatively homogenous groups (for a comprehensive description of the study see Williams, 2002). To this end a great variety of conditions researchers are likely to encounter in empirical applications has been analysed. Specifically, we address the following questions:

1. How do MECOSA's estimation methods perform with regard to the absolute goodness of parameter recovery?

2. What are the important factors which influence the relative goodness of parameter recovery?

To assess the absolute performance of MECOSA in situations where the researcher is unsure about the number of mixture components we recorded the number of times MECOSA failed to determine the correct number. In addition, we calculated the coverage rate of the nominal

95% interval for the parameter estimates. For the overall goodness-of-fit statistic we tested if the distributions in our simulation actually resemble the theoretically derived χ^2 distribution for large samples.

Relative goodness of parameter recovery is examined by calculating mean bias and mean variance. Since these are highly aggregated measures which might hide major differences at the level of individual parameters, ANOVA has been used to analyse which factors influence the individual parameter estimates.

3.1   Experimental Design

The experimental design of our study considerably extends the one underlying the Monte Carlo simulation on MECOSA published by Arminger, Stein & Wittenberg (1999) in Psychometrika. The single experimental factor "Estimation method" (MDE, EM, EMG) varied in their study has been augmented by five additional factors:

1. Distribution of exogenous variables (normal, skewed)

2. Number of groups (two, three)

3. Group proportions (equal, unequal (80:20 in the two groups case and 50:15:35 in the three groups case))

4. Separation of parameter values (far, close)

5. Knowledge about number of groups (known, unknown)

Even though Arminger, Stein & Wittenberg (1999) examined the effect of skewed exogenous regressors on the performance of MECOSA, a simulation study was only performed for the first stage of the MDE procedure and was based on a simple regression model. The results clearly showed that the conditional finite mixture approach in the case of non-normal regressors outperforms the unconditional approach. For this reason, the unconditional model was not further considered in our study. We nevertheless analysed if the distribution of the regressor variables has an influence on the quality of the MECOSA results in the context of a conditional structural model.

Since a mixture with two groups and equal group proportions as analysed in Arminger, Stein & Wittenberg (1999) represents a somewhat stylised situation we also explored cases with three groups and mixed group proportions. Given that both mixture components in the study by Arminger, Stein & Wittenberg (1999) were distinctly separated with respect to the population parameters of the underlying structural model, we additionally analysed groups where the parameters were closer together. Finally, we also considered situations where the true number of components is unknown and has to be estimated from the data. In total, this generates a six-way balanced factorial design with 96 unique experimental conditions (see Table 1). For each condition 500 valid data sets with 2000 observations each have been simulated with GAUSS (version 3.2.38).

		2 Groups				3 Groups			
		Equal		Unequal		Equal		Unequal	
		Skewed	Normal	Skewed	Normal	Skewed	Normal	Skewed	Normal
		Far Parameters							
	MDE	ASW							
Known	EM	ASW							
	EMG	ASW							
	MDE								
Unknown	EM								
	EMG								
		Close Parameters							
	MDE								
Known	EM								
	EMG								
	MDE								
Unknown	EM								
	EM Grad								

Note: The acronym ASW indicates the conditions examined in Arminger, Stein & Wittenberg (1999)

Table 1: Full Experimental Design of the Monte Carlo Simulation

3.2 Model for Data Generation

For reasons of comparability we selected the conditional structural equation model used by Arminger, Stein & Wittenberg (1999) to simulate the data. Their recursive model has two endogenous latent variables η_1 and η_2, each of which is measured by three continuous indicator variables (y_1, y_2, y_3 and y_4, y_5, y_6). The endogenous variables are both influenced by three exogenous regressor variables x_1, x_2 and x_3. In addition, the variable η_1 has a direct effect on the variable η_2. A diagram of the structural equation model is shown in Figure 1. Data are generated using group-specific parameterisations of this model. Apart from the factor loadings all parameters are varied across the components. For example, for the two components mixture the population values and the differences between the group parameters are reported in Table 2. The parameters used by (Arminger, Stein & Wittenberg, 1999) are designated as the "far" condition, Here, parameter differences are between 0.20 and 1.00 in absolute values. In contrast, for the "close" condition the corresponding values range from 0.00 to 0.30 with the exception of the intercepts (for the second group the intercepts in the measurement model are the same for "far" and "close" conditions).

Under the "normal" conditions all three exogenous variables are normally distributed with mean 0 and variance 1. For the "skewed" conditions the exogenous regressor variables consist of a standard Normal variable with mean 1 and variance 1 (x_1), a χ_1^2 variable with mean 1 and variance 2 (x_2), and a Bernoulli variable with mean 0.7 and variance 0.21 (x_3). The last two variables were standardised. The remaining error variables for which random values had to be generated are normally distributed.

θ_j	Group 1	Group 2			
		Far	Diff$_{12}$	Close	Diff$_{12}$
β_{21}	0.50	-0.50	1.00	0.20	0.30
γ_{11}	0.20	0.50	-0.30	0.10	0.10
γ_{12}	0.50	-0.50	1.00	0.40	0.10
γ_{13}	1.00	0.50	0.50	1.00	0.00
γ_{21}	-0.30	0.30	-0.60	-0.10	0.20
γ_{22}	0.00	1.00	-1.00	0.20	0.20
γ_{23}	0.50	-0.50	1.00	0.30	0.20
ψ_{11}	0.50	0.25	0.25	0.70	0.20
ψ_{22}	0.50	0.25	0.25	0.70	0.20
λ_{11}	1.00	1.00	0.00	1.00	0.00
λ_{21}	0.80	0.80	0.00	0.80	0.00
λ_{31}	0.70	0.70	0.00	0.70	0.00
λ_{42}	1.00	1.00	0.00	1.00	0.00
λ_{52}	0.90	0.90	0.00	0.90	0.00
λ_{62}	0.60	0.60	0.00	0.60	0.00
θ^{ϵ}_{11}	0.25	0.50	-0.25	0.30	-0.05
θ^{ϵ}_{22}	0.40	0.60	-0.20	0.50	-0.10
θ^{ϵ}_{33}	0.40	0.60	-0.20	0.50	-0.10
θ^{ϵ}_{44}	0.25	0.50	-0.25	0.30	-0.05
θ^{ϵ}_{55}	0.30	0.60	-0.30	0.40	-0.10
θ^{ϵ}_{66}	0.50	0.80	-0.30	0.50	0.00
ν_1	0.00	1.00	-1.00	1.00	-1.00
ν_2	0.00	1.00	-1.00	1.00	-1.00
ν_3	0.00	1.00	-1.00	1.00	-1.00
ν_4	0.00	1.00	-1.00	1.00	-1.00
ν_5	0.00	1.00	-1.00	1.00	-1.00
ν_6	0.00	1.00	-1.00	1.00	-1.00

Table 2: Population Parameters (2 groups)
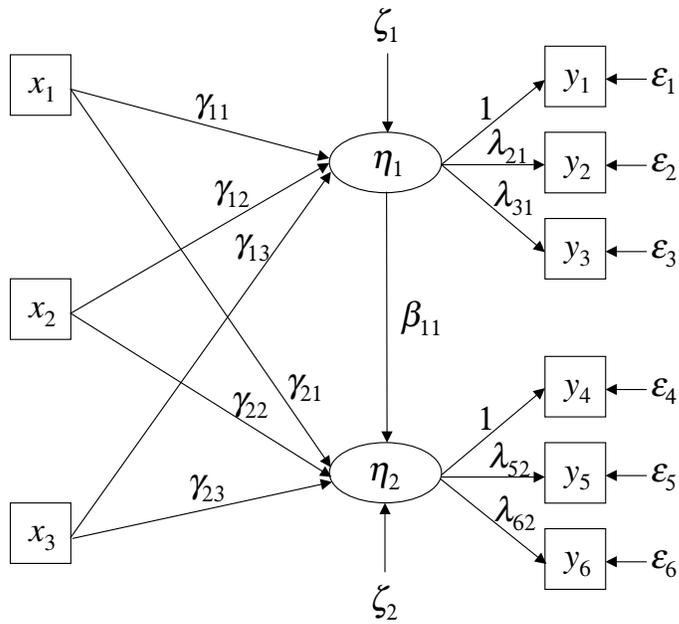
Figure 1: Structural Equation Model for Data Simulation

3.3 Results

The simulated data has been analysed using MECOSA 3.08 under Windows 98. The output files generated by the GAUSS simulation programme have been post-processed and analysed by the statistical computing software R.

The results we report are structured as follows:

- Comparison of our results with Arminger, Stein & Wittenberg (1999)

- Frequency of incorrectly estimated number of groups

- Coverage of the 95% confidence intervall

- Distribution of the MDE test statistic

- Mean bias

- Mean MAD

- Parameter-specific analysis of variance

3.3.1 Comparison With the Results of Arminger, Stein, and Wittenberg (1999)

We first compared our simulation results with those reported in (Arminger, Stein & Wittenberg, 1999). For the experimental conditions which were analysed in their study (see Table 1), Table 3 shows the means of the parameter estimates for both studies, calculated from 500 Monte Carlo replications each. In addition, the mean of the standard errors output by MECOSA and the standard deviations of the parameter estimates are given in parentheses. With respect to the means of the parameter estimates we obtained virtually identical results (differences occured only for the third decimal place). However, for the parameters γ_{13}, γ_{22} and γ_{23} some noticeable differences exist for the mean standard errors/standard deviations. Despite these deviations, overall our results agree well with those of (Arminger, Stein & Wittenberg, 1999). Since the means of the estimated parameters are almost identical with the true parameters, parameter recovery is quite good for all three estimation methods in the case of two groups with equal proportions and far parameters.

MDE

θ_j	True	First Component Arminger	This study	True	Second Component Arminger	This study
π_k	0.500	0.500 (0.017/0.011)	0.500 (0.015/0.010)	0.500	0.500 (0.000/0.011)	0.500 (0.000/0.010)
β_{21}	0.500	0.505 (0.052/0.052)	0.506 (0.051/0.051)	-0.500	-0.505 (0.094/0.100)	-0.500 (0.074/0.075)
γ_{11}	0.200	0.202 (0.030/0.031)	0.200 (0.028/0.028)	0.500	0.500 (0.027/0.026)	0.500 (0.024/0.024)
γ_{12}	0.500	0.501 (0.020/0.020)	0.500 (0.028/0.028)	-0.500	-0.500 (0.019/0.019)	-0.502 (0.023/0.022)
γ_{13}	1.000	0.996 (0.064/0.064)	1.000 (0.028/0.029)	0.500	0.500 (0.056/0.057)	0.501 (0.023/0.023)
γ_{21}	-0.300	-0.303 (0.032/0.034)	-0.304 (0.031/0.030)	0.300	0.303 (0.056/0.059)	0.300 (0.044/0.043)
γ_{22}	0.000	-0.004 (0.034/0.034)	-0.006 (0.040/0.038)	1.000	0.997 (0.053/0.055)	1.000 (0.045/0.046)
γ_{23}	0.500	0.500 (0.083/0.082)	0.497 (0.056/0.056)	-0.500	-0.495 (0.078/0.080)	-0.498 (0.045/0.047)
ψ_{11}	0.500	0.492 (0.037/0.037)	0.493 (0.035/0.036)	0.250	0.245 (0.030/0.031)	0.246 (0.026/0.026)
ψ_{22}	0.500	0.491 (0.038/0.039)	0.492 (0.035/0.035)	0.250	0.242 (0.036/0.038)	0.247 (0.027/0.029)

EM

θ_j	True	First Component Arminger	This study	True	Second Component Arminger	This study
π_k	0.500	0.500 (0.017/0.011)	0.501 (0.015/0.010)	0.500	0.500 (0.000/0.011)	0.499 (0.000/0.010)
β_{21}	0.500	0.505 (0.052/0.050)	0.500 (0.050/0.053)	-0.500	-0.506 (0.094/0.096)	-0.508 (0.074/0.074)
γ_{11}	0.200	0.202 (0.030/0.030)	0.201 (0.028/0.027)	0.500	0.500 (0.027/0.026)	0.501 (0.024/0.024)
γ_{12}	0.500	0.500 (0.020/0.019)	0.500 (0.028/0.027)	-0.500	-0.500 (0.019/0.019)	-0.501 (0.023/0.022)
γ_{13}	1.000	0.996 (0.064/0.062)	1.001 (0.028/0.029)	0.500	0.500 (0.056/0.055)	0.503 (0.023/0.023)
γ_{21}	-0.300	-0.302 (0.032/0.033)	-0.300 (0.031/0.031)	0.300	0.303 (0.056/0.057)	0.303 (0.045/0.044)
γ_{22}	0.000	-0.003 (0.034/0.033)	0.001 (0.039/0.041)	1.000	0.997 (0.052/0.054)	0.995 (0.046/0.045)
γ_{23}	0.500	0.497 (0.083/0.080)	0.499 (0.055/0.057)	-0.500	-0.495 (0.078/0.078)	-0.496 (0.046/0.046)
ψ_{11}	0.500	0.496 (0.037/0.036)	0.498 (0.035/0.034)	0.250	0.246 (0.030/0.030)	0.246 (0.025/0.025)
ψ_{22}	0.500	0.495 (0.038/0.038)	0.493 (0.035/0.034)	0.250	0.244 (0.035/0.037)	0.246 (0.027/0.025)

EMG

θ_j	True	First Component Arminger	This study	True	Second Component Arminger	This study
π_k	0.500	0.500 (0.017/0.011)	0.499 (0.015/0.010)	0.500	0.500 (0.000/0.011)	0.500 (0.000/0.010)
β_{21}	0.500	0.505 (0.052/0.050)	0.501 (0.051/0.051)	-0.500	-0.507 (0.094/0.096)	-0.503 (0.073/0.073)
γ_{11}	0.200	0.202 (0.030/0.030)	0.202 (0.029/0.029)	0.500	0.500 (0.027/0.026)	0.500 (0.024/0.022)
γ_{12}	0.500	0.500 (0.020/0.019)	0.503 (0.028/0.026)	-0.500	-0.500 (0.019/0.019)	-0.500 (0.023/0.023)
γ_{13}	1.000	0.995 (0.064/0.062)	1.004 (0.028/0.029)	0.500	0.500 (0.056/0.055)	0.500 (0.023/0.022)
γ_{21}	-0.300	-0.302 (0.032/0.033)	-0.300 (0.031/0.031)	0.300	0.303 (0.056/0.057)	0.302 (0.044/0.042)
γ_{22}	0.000	-0.003 (0.034/0.033)	0.001 (0.039/0.038)	1.000	0.996 (0.052/0.054)	0.998 (0.045/0.045)
γ_{23}	0.500	0.496 (0.083/0.080)	0.500 (0.056/0.054)	-0.500	-0.495 (0.078/0.078)	-0.498 (0.045/0.047)
ψ_{11}	0.500	0.496 (0.037/0.036)	0.496 (0.035/0.035)	0.250	0.246 (0.030/0.030)	0.250 (0.026/0.026)
ψ_{22}	0.500	0.495 (0.038/0.038)	0.498 (0.036/0.035)	0.250	0.243 (0.035/0.037)	0.247 (0.026/0.026)

Note: The first values in parentheses are the mean standard errors of 500 replications and the second values are the standard deviations of the parameter estimates caculated from the replications

Table 3: Comparison With Arminger, Stein, and Wittenberg (1999)

3.3.2 Estimating the Number of Components

The determination of the "correct" number of components in the data is a crucial step in finite mixture analysis since it has a major impact on the validity of the parameter estimates. For all experimental conditions with an unknown number of groups we used the modified likelihood ratio test (ad hoc test) implemented in MECOSA as the default option to estimate the number of groups (see Section 2.2.3). The simulation programme was set up such that if MECOSA identified an incorrect number of groups for a specific sample the results of that replication would be discarded and the iteration started again. This procedure secured that we achieved 500 valid data sets, which is a prerequisite for meaningful comparisons between results for different conditions (Paxton et al. 2001). Whereas for all "two groups" conditions and the "three groups" conditions with equal group proportions the number of components was almost always estimated correctly, for those "three groups" conditions with close parameters and mixed proportions we had to simulate between 1669 and 2197 samples to achieve the required number of 500 valid samples (see Table 4). Since the time needed to let the parametric bootstrap test detect the number of components for at least 500 replications in 96 conditions was prohibitive, we focused on a specific condition to compare its performance with that of the ad hoc test. Using the MDE procedure, the failure rate of the parametric bootstrap test under the condition of three components, mixed proportions, close parameters and normal regressors was 16.7% compared to 76.5% for the ad hoc test. Although these results come from only one condition and thus need further investigations, the differences are considerable. Thus, on the basis of our results it is recommended that researchers use the parametric bootstrap test (and alternative measures like BIC) whenever possible to estimate the number of components in their data.

		Two Groups		Three Groups	
		Equal	Uneq.	Equal	Uneq.
		Close Parameters			
	MDE	0	1	1	1697
Skewed	EM	0	0	3	1532
	EMG	0	1	3	1691
	MDE	0	0	3	1631
Normal	EM	1	0	3	1169
	EMG	0	0	4	1260
		Far Parameters			
	MDE	3	0	2	4
Skewed	EM	0	1	2	2
	EMG	3	0	2	4
	MDE	0	0	0	17
Normal	EM	2	3	0	19
	EMG	1	2	0	15

Table 4: Frequency of Incorrectly Estimated Number of Groups

3.3.3 Absolute Goodness of Estimation

Coverage of the Nominal 95% Confidence Interval   Coverage was estimated by calculating the proportion of replications for which the 95% confidence interval contains the true parameter value. The results are summarised in Table 5. For "far parameters" 95% coverage is quite close to the expected nominal value under almost all conditions. With coverage around 90% MDE, however, does not perform as good as the other estimation procedures in the three groups cases. A similar picture emerges for "close parameters". Although compared to the "far" conditions results for the two groups cases as well as for the three groups conditions with equal group proportions are slightly worse, coverage is still rather good. A significant drop in coverage occurs for unequal group proportions, especially with respect to the MDE procedure where coverage is only about 82%. Thus, whereas for the two groups conditions the choice of the estimation procedure has little effect, it becomes relevant for "close" parameters when the three groups proportions are mixed.

		Two Groups				Three Groups	
		Known		Unknown		K & U[a]	
		Equal	Uneq.	Equal	Uneq.	Equal	Uneq.
		Close Parameters					
	MDE	90.9	90.9	92.3	90.4	90.3	82.1
Skewed	EM	93.6	92.6	93.2	93.5	93.8	90.0
	EMG	91.5	90.9	91.5	91.5	93.5	90.2
	MDE	90.9	90.2	92.3	90.5	88.2	82.3
Normal	EM	93.3	93.1	91.0	92.9	91.0	89.8
	EMG	91.4	91.3	93.0	92.7	93.4	90.1
		Far Parameters					
	MDE	92.2	93.9	94.5	94.2	89.8	88.8
Skewed	EM	93.2	95.1	93.2	93.3	94.7	94.0
	EMG	95.2	93.3	93.1	93.2	94.6	94.2
	MDE	94.5	94.2	94.3	94.1	90.3	90.3
Normal	EM	95.3	93.0	95.2	93.6	94.8	93.7
	EMG	93.2	94.9	93.1	93.3	94.5	94.4

[a]Known and unknown conditions are merged

Table 5: Coverage of the Nominal 95% Confidence Interval

Distribution of Test Statistic   Under the null hypothesis that the hypothesised component-specific models are the true models the MDE test statistic $Q(\hat{\vartheta})$ follows asymptotically a central χ^2 distribution with $r - d$ degrees of freedom. In order to assess if this holds true for the different experimental conditions in our simulation the values of $Q(\hat{\vartheta})$ were recorded for each of the 500 replications. We then compared the resulting distributions with the appropriate χ^2 distribution (with 44 degrees of freedom in the "two groups" case) using a Kolmogorov-Smirnov test. Table 6 summarises the test results only for the two components mixtures and Figure 2 shows both the theoretical and empirical distributions for the "unknown-2 groups" conditions. We do not report the results for the three components simulation data because

the p values of the Kolmogorov-Smirnov test were all less than 0.0000, which means that the corresponding observed distribution of the test statistic does not follow a χ^2 distribution. Even for the "two groups" conditions several tests yielded significant p values ($\alpha = .05$), but no clear pattern emerged. Overall, the results suggest that the theoretical behaviour of $Q(\hat{\vartheta})$ is not well reproduced even in data sets with a rather large number of observations.

	Known		Unknown	
	Equal	Uneq.	Equal	Uneq.
	Close Parameters			
Skewed	0.00016	0.51821	0.00209	0.10759
Normal	0.03201	0.05199	0.00027	0.01669
	Far Parameters			
Skewed	0.00415	0.04312	0.03549	0.00000
Normal	0.91119	0.00012	0.00359	0.00000

Table 6: p Values of Kolmogorov-Smirnov Test on χ^2 Distribution of $Q(\hat{\vartheta})$ (2 Groups)

3.3.4   Relative Goodness of Estimation

To evaluate the relative performance of the different estimation procedures in MECOSA we use measures of estimator bias and variability. Since the experimental design resulted in 96 different conditions it is not possible to report these measures for every parameter. Instead, for each condition we aggregate over all parameters and replications although we note that we hereby risk to lose some important information. For example, a high mean bias for a specific condition might be the result of having only a few parameters showing extremely high biases or due to all parameters having moderately high biases.

Mean Bias   The fundamental measure of estimator bias for a single parameter is $E[\hat{\theta}] - \theta$. The magnitude of this measure among other things obviously depends on the size of θ. Dividing this expression by the true parameter θ would eliminate the scale effect, but this relative bias is only defined for non-zero population values. Since there are a few zero population parameters in our study, we define our measure of mean bias as follows:

$$\text{mean } B = \frac{1}{n_\theta} \sum_{j=1}^{n_\theta} \left[\left(\frac{1}{n_r} \sum_{i=1}^{n_r} \hat{\theta}_{ij} \right) - \theta_j \right] \tag{10}$$

with $i = 1, \ldots, n_r$, where n_r is the number of replications (here, 500 in all conditions), and $j = 1, \ldots, n_\theta$, where n_θ is the number of free parameters (here, for example, 48 in all conditions of the two components mixtures). $\hat{\theta}_{ij}$ denotes the ith estimate of parameter j and θ_j its true value. Although the "close" and "far" conditions use different parameter sets, it can reasonably be assumed that the scale effect is so small that comparisons of the estimator bias across these conditions are admissible.

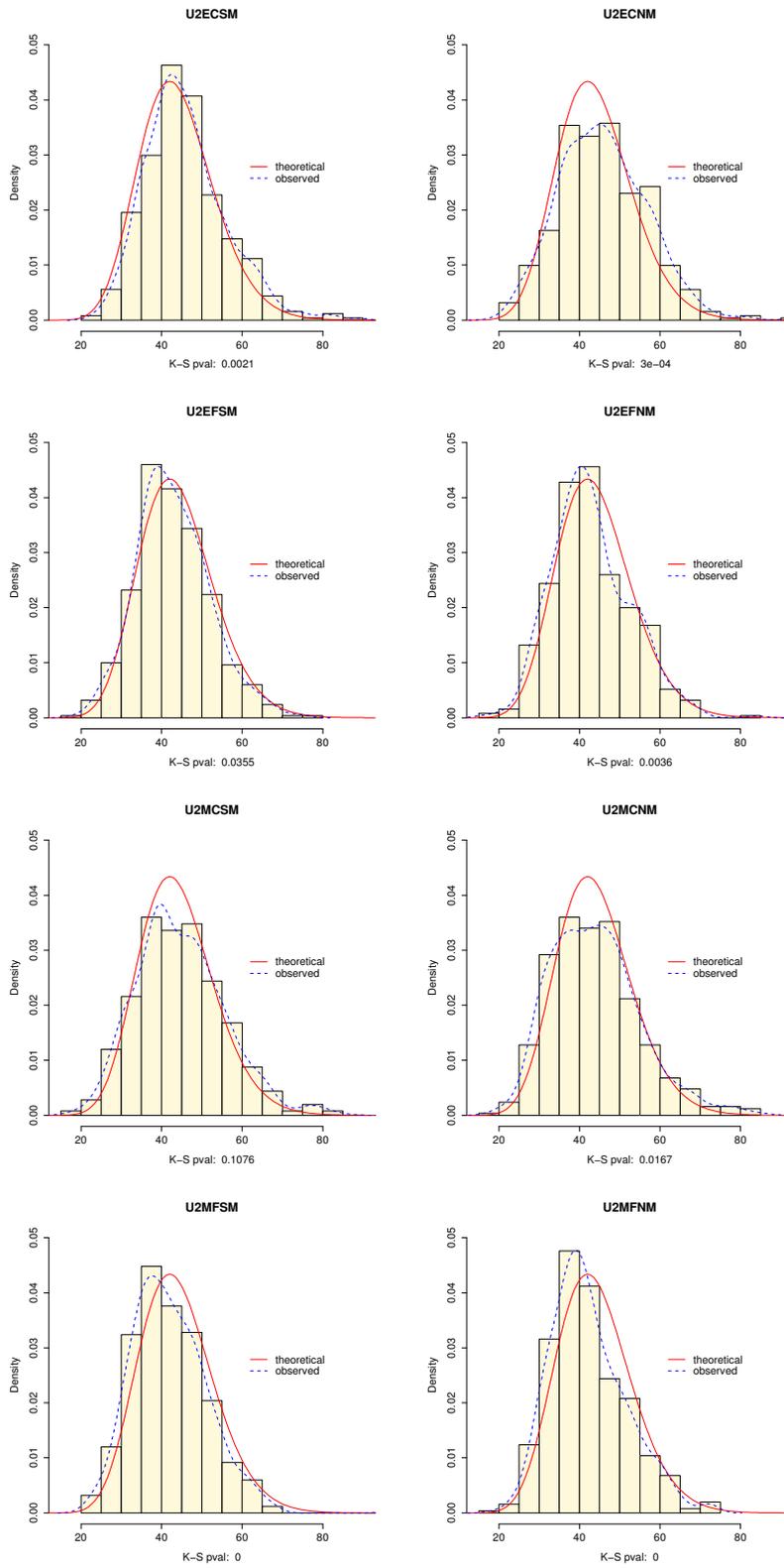The mean biases for the estimation of two components mixtures are shown in Table 7. First of

15

Figure 2: Distribution of $Q(\hat{\boldsymbol{\vartheta}})$ (Unknown group number, 2 Groups)

		Known		Unknown	
		Equal	Uneq.	Equal	Uneq.
		Close Parameters			
Skewed	MDE	-0.0049	-0.0116	-0.0049	-0.0127
	EM	-0.0010	-0.0018	-0.0014	-0.0012
	EMG	-0.0021	-0.0018	-0.0013	-0.0020
Normal	MDE	-0.0047	-0.0134	-0.0061	-0.0118
	EM	-0.0015	-0.0020	-0.0020	-0.0016
	EMG	-0.0018	-0.0019	-0.0029	-0.0033
		Far Parameters			
Skewed	MDE	-0.0016	-0.0033	-0.0022	-0.0028
	EM	-0.0003	-0.0004	-0.0002	-0.0007
	EMG	-0.0004	-0.0006	-0.0007	-0.0006
Normal	MDE	-0.0018	-0.0026	-0.0019	-0.0024
	EM	-0.0007	-0.0010	-0.0004	-0.0004
	EMG	-0.0002	-0.0005	-0.0005	-0.0006

Table 7: Mean Bias for Each Condition (2 Groups)

all, we notice that the mean biases are all negative (although some biases for individual parameters were positive). This means that on average the estimates are less than the true population values. Compared to the EM and EMG algorithms, using the MDE procedure consistently produces a considerably higher mean bias. For the two other estimation procedures only minor differences occur and none seems to have an advantage over the other. Overall, the "close" conditions lead to higher biases than the "far" conditions, where especially high biases occur for the "MDE-close-unequal" conditions. Given that only those replications were included where MECOSA correctly estimated the number of groups it is plausible that "known" and "unknown" conditions lead to similar results.

Mean MAD   Estimator variability can be determined, for example, by calculating the mean squared error or the mean MAD (mean absolute deviation). Since in our study both measures lead to similar interpretations with respect to the relative performance of the estimation procedures under the different conditions, we exclusively focus on the mean MAD:

$$\text{mean MAD} = \frac{1}{n_\theta} \sum_{j=1}^{n_\theta} \left[\frac{1}{n_r} \sum_{i=1}^{n_r} \left| \hat{\theta}_{ij} - \theta_j \right| \right] \tag{11}$$

The mean MAD's for all "two groups" conditions are shown in Table 8. The results show that the MDE procedure consistently produces the highest variance although the distance to the EM and EMG biases is only modest. This is especially true for the "far" conditions. For EM and EMG again only negligible differences occur. A pronounced effect on the variance results from the separation of the population parameters. Here, the "close" conditions have a considerably higher variance. Overall, the highest variance is reported for the "closed-mixed" treatments in combination with the MDE procedure.

17

		Known		Unknown	
		Equal	Uneq.	Equal	Uneq.
		Close Parameters			
	MDE	0.0339	0.0443	0.0338	0.0452
Skewed	EM	0.0316	0.0409	0.0319	0.0400
	EMG	0.0317	0.0399	0.0317	0.0392
	MDE	0.0337	0.0456	0.0340	0.0451
Normal	EM	0.0320	0.0400	0.0322	0.0396
	EMG	0.0323	0.0398	0.0319	0.0402
		Far Parameters			
	MDE	0.0230	0.0283	0.0227	0.0282
Skewed	EM	0.0221	0.0272	0.0226	0.0268
	EMG	0.0223	0.0270	0.0224	0.0272
	MDE	0.0222	0.0269	0.0223	0.0271
Normal	EM	0.0216	0.0265	0.0216	0.0259
	EMG	0.0215	0.0265	0.0217	0.0257

Table 8: Mean MAD for Each Condition (2 Groups)

To sum up, it can be noticed that the two-stage MDE procedure leads both to the highest mean bias and variance. This applies especially to conditions with unequal group proportions and close parameters. A possible explanation for the poor parameter recovery in this situations is that the group proportions were not estimated well. In this case, a large number of observations is assigned to the wrong group, leading to a high variance in the components and degrading the chance of retrieving the correct parameters.

3.3.5 Parameter-Specific Analysis of Variance

Because mean bias and mean MAD are highly aggregated measures of the quality of an estimation procedure, we additionally explored which factors influence parameter estimates at the level of individual parameters by performing analysis of variance (ANOVA). The dependent variable contains the group-specific parameter estimate for each replication in each condition, which results in a balanced ANOVA with 500 replications in each cell. The independent variables were the four binary variables denoting (1) known or unknown number of groups, (2) equal or unequal group proportions, (3) close or far parameters and (4) skewed or normally distributed regressors. Here, one indicates the attribute level mentioned first (e.g., known = 1) and zero the alternative level (e.g., unknown = 0). To determine the influence of the estimation method we additionally employed dummy coding: Two exogenous dummy variables were created indicating whether EM or EM Gradient estimation was used (= 1) or not used (= 0). Thus using MDE is the baseline condition. Since for the first group only a single set of population values has been used for simulation, the dummy variable "close" was not included in the ANOVA of parameter estimates for this group. Regarding the second group, the interesting question is if there are some significant effects on the parameter estimates after controlling for the dominant influence of the variable "close".

18

θ_j	Intercept	Known	Equal	Close	Skewed	EM	EM Grad	Adj. R^2
$\beta_{21}^{(2)}$	−0.4912		−0.0056	0.6939				0.93
$\gamma_{11}^{(2)}$	0.5032			−0.3991		−0.0027		0.96
$\gamma_{12}^{(2)}$	−0.5003			0.9015				0.99
$\gamma_{13}^{(2)}$	0.5009	−0.0022		0.4985				0.97
$\gamma_{21}^{(2)}$	0.2952		0.0037	−0.3996				0.92
$\gamma_{22}^{(2)}$	1.0037			−0.8027				0.98
$\gamma_{23}^{(2)}$	−0.4999			0.8082				0.95
$\psi_{11}^{(2)}$	0.2373		0.0081	0.4360		0.0072	0.0069	0.91
$\psi_{22}^{(2)}$	0.2371		0.0114	0.4332		0.0066	0.0050	0.92
$\theta_{11}^{\epsilon,(2)}$	0.2856		0.0054	0.1962		0.0127	0.0122	0.80
$\theta_{22}^{\epsilon,(2)}$	0.2844		0.0073	0.1957		0.0124	0.0129	0.49
$\theta_{33}^{\epsilon,(2)}$	0.3809		0.0078	0.1941		0.0161	0.0165	0.53
$\theta_{44}^{\epsilon,(2)}$	0.4770		0.0102	0.2931		0.0174	0.0170	0.79
$\theta_{55}^{\epsilon,(2)}$	0.7988		0.2996	0.0015				0.79
$\theta_{66}^{\epsilon,(2)}$	0.2012		0.2993	0.0014				0.88
			p values					
β_{21}^2	0.00	0.59	0.00	0.00	0.50	0.05	0.04	
γ_{11}^2	0.00	0.74	0.20	0.00	0.06	0.00	0.10	
γ_{12}^2	0.00	0.48	0.75	0.00	0.56	0.45	0.23	
γ_{13}^2	0.00	0.00	0.13	0.00	0.10	0.68	0.08	
γ_{21}^2	0.00	0.76	0.00	0.00	0.74	0.03	0.18	
γ_{22}^2	0.00	0.46	0.55	0.00	0.02	0.29	0.99	
γ_{23}^2	0.00	0.77	0.17	0.00	0.25	0.08	0.08	
ψ_{11}^2	0.00	0.33	0.00	0.00	0.75	0.00	0.00	
ψ_{22}^2	0.00	0.05	0.00	0.00	0.80	0.00	0.00	
$\theta_{11}^{\epsilon,(2)}$	0.00	0.10	0.00	0.00	0.72	0.00	0.00	
$\theta_{22}^{\epsilon,(2)}$	0.00	0.92	0.00	0.00	0.70	0.00	0.00	
$\theta_{33}^{\epsilon,(2)}$	0.00	0.86	0.00	0.00	0.97	0.00	0.00	
$\theta_{44}^{\epsilon,(2)}$	0.00	0.49	0.00	0.00	0.91	0.00	0.00	
$\theta_{55}^{\epsilon,(2)}$	0.00	0.41	0.00	0.00	0.32	0.21	0.01	
$\theta_{66}^{\epsilon,(2)}$	0.00	0.40	0.00	0.00	0.32	0.22	0.01	

Table 9: ANOVA Results for the Main Effects Model (2 Groups)

The results for a simple main effects model with no interaction effects are given in Table 9. The parameter superscript contains the group number in parentheses and the subscript indicates the position of the parameter in the corresponding LISREL matrix. An entry in the table means that the variable has an effect at the 1% level of significance (see the p values in the lower half of the table). ANOVA results with an adjusted R^2 less than 10% are not reported.

As can be seen from the parameters included in the first column of Table 9, the most striking result is that estimates for group one are not affected (besides the fact that all ANOVA's had an adjusted R^2 lower than 10%, the important aspect is that although there were a few significant coefficients no clear pattern emerged). This is most probably due to the sample size and misclassification problems already discussed before. In contrast, estimates for the parameters in group two are consistently affected by some experimental conditions. Given that the true parameters for group two belong either to the "close" or "far" condition it is obvious that every parameter is influenced by the variable "close". After controlling for this factor, some further significant effects on the parameter estimates, especially the error variances, exist. The variable "equal" as well as the estimation methods used have an influence. However, the parameter estimates are not affected by the distribution of the regressor variables and the

θ_j	Intercept	Equal	Close	Equal $\times$ Close	EM	EM Grad	Adj. R^2
$\beta_{21}^{(2)}$	-0.493		0.761	-0.065	-0.023	-0.016	0.93
$\gamma_{11}^{(2)}$	0.502		-0.390	-0.008			0.96
$\gamma_{12}^{(2)}$	-0.499		0.908	-0.008	-0.003	-0.003	0.99
$\gamma_{13}^{(2)}$	0.500		0.496	0.004			0.97
$\gamma_{21}^{(2)}$	0.299		-0.420	0.019	0.008	0.006	0.93
$\gamma_{22}^{(2)}$	0.992		-0.826	0.028	0.011	0.006	0.97
$\gamma_{23}^{(2)}$	-0.503		0.780	0.026	0.007	0.005	0.95
$\psi_{11}^{(2)}$	0.245	0.005	0.470	-0.030	-0.006		0.92
$\psi_{22}^{(2)}$	0.242	0.006	0.449	-0.007	-0.004		0.93
$\theta_{11}^{\epsilon,(2)}$	0.286		0.170	0.029	0.016	0.012	0.79
$\theta_{22}^{\epsilon,(2)}$	0.480	0.006	0.089	0.009	0.016	0.014	0.53
$\theta_{33}^{\epsilon,(2)}$	0.482	0.005	0.086	0.014	0.016	0.013	0.51
$\theta_{44}^{\epsilon,(2)}$	0.286	0.004	0.173	0.026	0.016	0.011	0.78
$\theta_{55}^{\epsilon,(2)}$	0.382	0.006	0.174	0.023	0.018	0.013	0.79
$\theta_{66}^{\epsilon,(2)}$	0.478	0.006	0.281	0.016	0.020	0.017	0.89
$\nu_1^{(2)}$	0.979		-0.091	0.090	0.038	0.026	0.24
$\nu_2^{(2)}$	0.981		-0.092	0.091	0.036	0.025	0.25
$\nu_3^{(2)}$	0.980		-0.090	0.087	0.038	0.025	0.26
$\nu_4^{(2)}$	0.980		-0.084	0.086	0.034	0.022	0.22
$\nu_5^{(2)}$	0.978		-0.085	0.086	0.035	0.024	0.23
$\nu_6^{(2)}$	0.982		-0.086	0.084	0.032	0.022	0.24
π_1	0.793	-0.300	-0.038	0.042	0.012	0.008	0.96
π_2	0.207	0.300	0.039	-0.042	-0.012	-0.008	0.96
p values							
$\beta_{21}^{(2)}$	0.000	0.019	0.000	0.000	0.000	0.000	
$\gamma_{11}^{(2)}$	0.000	0.351	0.000	0.000	0.024	0.069	
$\gamma_{12}^{(2)}$	0.000	0.193	0.000	0.000	0.001	0.005	
$\gamma_{13}^{(2)}$	0.000	0.871	0.000	0.009	0.030	0.271	
$\gamma_{21}^{(2)}$	0.000	0.033	0.000	0.000	0.000	0.000	
$\gamma_{22}^{(2)}$	0.000	0.127	0.000	0.000	0.000	0.000	
$\gamma_{23}^{(2)}$	0.000	0.644	0.000	0.000	0.000	0.011	
$\psi_{11}^{(2)}$	0.000	0.006	0.000	0.000	0.000	0.625	
$\psi_{22}^{(2)}$	0.000	0.000	0.000	0.001	0.008	0.464	
$\theta_{11}^{\epsilon,(2)}$	0.000	0.064	0.000	0.000	0.000	0.000	
$\theta_{22}^{\epsilon,(2)}$	0.000	0.000	0.000	0.000	0.000	0.000	
$\theta_{33}^{\epsilon,(2)}$	0.000	0.000	0.000	0.000	0.000	0.000	
$\theta_{44}^{\epsilon,(2)}$	0.000	0.006	0.000	0.000	0.000	0.000	
$\theta_{55}^{\epsilon,(2)}$	0.000	0.000	0.000	0.000	0.000	0.000	
$\theta_{66}^{\epsilon,(2)}$	0.000	0.000	0.000	0.000	0.000	0.000	
$\nu_1^{(2)}$	0.000	0.655	0.000	0.000	0.000	0.000	
$\nu_2^{(2)}$	0.000	0.787	0.000	0.000	0.000	0.000	
$\nu_3^{(2)}$	0.000	0.775	0.000	0.000	0.000	0.000	
$\nu_4^{(2)}$	0.000	0.785	0.000	0.000	0.000	0.000	
$\nu_5^{(2)}$	0.000	0.759	0.000	0.000	0.000	0.000	
$\nu_6^{(2)}$	0.000	0.669	0.000	0.000	0.000	0.000	
π_1	0.000	0.000	0.000	0.000	0.000	0.000	
π_2	0.000	0.000	0.000	0.000	0.000	0.000	

Table 10: ANOVA Results for the Interaction Model (2 Groups)

knowledge of the true number of groups. The latter result confirms the finding of (Arminger, Stein & Wittenberg, 1999) that parameter recovery was not negatively affected by skewed regressors in the context of a conditional finite mixture model.

Table 10 reports the significant coefficients and p values for an additional model with an "Equal $\times$ Close" interaction term and no terms for the variables "known" and "skewed" as they were not significant in the main effects model. Although the factors "group proportions" and "separation of parameter values" where orthogonal in our simulation design the interaction term has a consistently significant effect on the parameter estimates. Compared to the main effects model the estimation method has a significant effect on a much larger number of parameters. This result is in line with the finding that MDE in conjunction with mixed proportions and close parameters had a quite pronounced effect on mean bias and variance.

4   Summary and Conclusions

In this study we report the results of an extensive Monte Carlo simulation analysing the performance of the MECOSA approach to conditional finite mixture structural equation models when unobserved heterogeneity exists in a population. Here, we assume that a heterogenous population consists of an unknown, but limited number of relatively homogenous subpopulations. Specifically, we tested the three estimation methods (MDE, EM and EMG) available in MECOSA and examined which experimental factors had a substantial influence on their (relative) performance.

We first compared our results with those reported in Arminger, Stein & Wittenberg (1999) and found that both were in close correspondence. Thus, for two groups with equal proportions and a clear separation of the parameter sets, the parameters used for simulations were very well recovered by all three estimation methods.

Since in empirical applications of the finite mixture approach the true number of components is typically unknown to the researcher, he/she has to rely on statistical criteria to determine the number of groups. The findings of our study clearly show that the modified likelihood ratio test (ad hoc test) is inappropriate when the sample consists of three groups with unequal mixing proportions and parameters are rather close together. Because such a situation does not seem to be unrealistic for empirical applications, we thus recommend to use other criteria, for example the parametric bootstrap test which is also implemented in MECOSA, to determine the number of components.

Overall, MECOSA performed best when the group proportions where equal and parameters where distinctly separated. Even in this situation both EM algorithms (EM and EMG) which directly estimate the fundamental parameters performed slightly better with regard to mean bias and variability than the two-stage MDE procedure. MECOSA performed worst under conditions of close parameters and mixed proportions. As an ANOVA at the level of individual parameters has shown, a clear interaction effect between both factors exist. The results also made clear that the MDE procedure performs considerably worse under these conditions.

21

Since these are situations which typically occur in empirical studies (mixed proportions and/or close parameters), we suggest to use only the first stage of the MDE procedure to test for the number of components and to estimate the fundamental parameters by using the EMG method. Such a strategy was already proposed by Arminger, Stein & Wittenberg (1999) albeit for very large models.

Although our simulation design was set up such that various conditions researchers are likely to encounter in their empirical applications were covered, it inevitably is subject to a whole lot of restrictions. For example, sample size ($n = 2000$) has not been varied in our study. Thus, future research should analyse the effect different sample sizes have on the performance of MECOSA. Also the structural model used for data simulation was rather small for which reason more complex models should be considered. Further investigations are also needed concerning alternative overall test statistics (e.g., RMSEA). Because the ad hoc test used to estimate the number of mixture components performed badly under specific conditions, a comprehensive simulation study on alternative criteria is essential. Finally, an interesting aspect for further studies concerns the robustness of MECOSA against violations of the assumption that the endogenous variables are normally distributed conditional on the exogenous regressors.

References

Aitkin, M., Anderson, D., and Hinde, J. (1981). Statistical modeling of data on teaching styles. *Journal of the Royal Statistical Society A 144*, 419–461.

Aitkin, M. and Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society B 47*, 67–75.

Allenby, G. M. and Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics 89*(4), 57–78.

Ansari, A., Jedidi, K., and Jagpal, H. S. (2000). A hierarchical bayesian methodology for treating heterogenity in structural equation models. *Marketing Science 19*(4), 328–347.

Arminger, G. and Stein, P. (1997). Finite mixtures of covariance structure models with regressors. *Sociological Methods & Research 26*(2), 148–182.

Arminger, G., Stein, P., and Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika 64*(4), 475–494.

Arminger, G., Wittenberg, J., and Schepers, A. (1996). *MECOSA 3 User Guide*. Friedrichsdorf/Ts.: ADDITIVE GmbH.

Bagozzi, R. P., Yi, Y., and Nassen, K. D. (1999). Representation of measurement error in marketing variables: Review of approaches and extension to three-facet designs. *Journal of Econometrics 89*(1–2), 393–421.

Blåfield, E. (1980). Clustering of observations from finite mixtures with structural information. Jyväsyklä Studies in Computer Science, Economics & Statistics 2, Jyväsyklä University, Jyväsyklä, Finland.

Bollen, K. A. (1989). *Structural Equation Modeling with Latent Variables*. New York: Wiley.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B 39*, 1–38.

Dolan, C. V. and van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika 63*(3), 227–253.

Everitt, B. S. (1981). A Monte Carlo investigation of the likelihood ratio test for the number of componenents in a mixture of normal distributions. *Multivariate Behavioural Research 16*, 171–180.

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions*. London: Chapman and Hall.

Feng, Z. D. and McCulloch, C. E. (1996). Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society B 58*(3), 609–617.

Görz, N., Hildebrandt, L., and Annacker, D. (2000). Analyzing multigroup data with structural equation models. SFB 373 Discussion Paper, No. 11, Humboldt University, Berlin, Germany.

Jagpal, S. (1999). *Marketing Strategy and Uncertainty*. Oxford: Oxford University.

Jedidi, K., Jagpal, H. S., and DeSarbo, W. S. (1997a). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science 16*(1), 39–59.

Jedidi, K., Jagpal, H. S., and DeSarbo, W. S. (1997b). STEMM: A general finite mixture structural equation model. *Journal of Classification 14*, 23–50.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics 36*, 318–324.

McLachlan, G. J. and Basford, K. E. (1988). *Mixture Models*. New York: Marcel Dekker.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.

Muthén, B. O. and Muthén, L. K. (1998). *Mplus User's Guide*. Los Angeles: Muthén & Muthén.

Muthén, B. O. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics 55*(2), 463–469.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., and Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modelling 8*(2), 287–312.

Wedel, M., Kamakura, W., and Böckenholt, U. (2000). Marketing data, models and decisions. *International Journal of Research in Marketing 17*(2–3), 203–208.

Williams, J. (2002). *Mean and Covariance Structure Models for Finite Mixtures*. Ph. D. thesis, Department of Marketing, Otago University, Dunedin, New Zealand.

Wolfe, J. H. (1971). A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions. Technical Bulletin STB 72-2, Naval Personnel and Training Research Laboratory, San Diego, USA.

Yung, Y.-F. (1994). *Finite Mixtures in Confirmatory Factor-Analytic Models*. Ph. D. thesis, Department of Psychology, University of California, Los Angeles, USA.

Yung, Y.-F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika 62*(3), 297–330.