Distance models for analysis of multivariate binary data

Worku, H.M.

Author: Worku, H.M.
Title: Distance models for analysis of multivariate binary data
Issue Date: 2018-12-20

Chapter 2

Effects of a Small Number of Dichotomous Indicators in Latent Variable Modeling: A Simulation Study

Abstract

Structural equation models were originally proposed for the analysis of continuous or interval indicator variables. Recently, factor analysis and structural equation models have been applied for data with dichotomous indicators and with only a few indicators per latent variable, i.e. 2 or 3. We investigated the performance of Confirmatory Factor Analysis (CFA) and the Multiple Indicators MultIple Causes (MIMIC) model for dichotomous indicators in comparison with interval indicators in a Monte Carlo simulation study.

The performance of both CFA and the MIMIC model was studied in terms of the quality of recovering the true factor scores and the incidence of improper solutions, more specifically non-convergence and *Heywood* cases. Furthermore, in the case of the MIMIC model, the focus was on the type-I error rate and power.

We showed that both CFA and the MIMIC model performed poorly with a small number of dichotomous indicators, i.e., (1) improper solutions occurred much more frequently; (2) the true factor scores are poorly recovered; (3) the type-I error rates are too conservative mostly and inflated sometimes; and (4) the observed power was weak.

2.1 Introduction

Latent variable models are a general class of models that are used for analyzing multivariate data (Bartholomew & Knott, 1999; Skrondal & Rabe-Hesketh, 2004). In Latent Variable (LV) models the multivariate observed variables (manifest variables) are treated as dependent variable, and one or more unobserved variables (latent variables) are treated as independent variables. The observed variables are also known as indicators because they are used as an indirect measure of the latent variables.

The Latent variables can be interval or categorical. As displayed in Table 2.1, there are four classes of LV models based on the cross-classification of whether the observed variable and/or latent variable is interval and/or categorical (Bartholomew & Knott, 1999). Our main focus in this paper will be on Confirmatory Factor Analysis (CFA), which is a special case of Structural Equation Modeling (SEM: Thurstone, 1947; Jöreskog & Sörbom, 1981; Christoffersson, 1975; B. Muthen, 1978; Bock & Lieberman, 1970; Mislevy, 1986). Tomarken and Waller (2005) provided a detailed literature review on Structural Equation Modeling (SEM) focusing on its strengths, limitations, and misconceptions.

Table 2.1: Classes of Latent Variable Models.

	Latent variable	
Observed variable	Interval	Categorical
Interval	Factor Analysis/ Structural Equation Modeling	Latent Profile Analysis/ Mixture Modeling
Categorical	Item Response Theory/ Latent Trait Analysis	Latent Class Analysis

Traditionally SEM focuses on the analysis of continuous (or interval) indicator variables. Many studies have been performed to investigate the performance of structural equation models (Boomsma, 1983, 1985; J. C. Anderson & Gerbing, 1984; Acito & Anderson, 1986). Recently, in clinical psychological research structural equation models have been proposed for the analysis of comorbidity of depressive and anxiety disorders (Krueger, 1999; Beesdo-Baum et al., 2009). A typical characteristic of these models is that the indicators are dichotomous, i.e the indicators indicate whether someone has or does not have a particular disorder, and that there are only a few indicators per latent variable, i.e. 2 or 3. We believe the application of structural equation models in such a scenario (i.e., dichotomous indicators with a few number of variables per factor) is not adequate enough to obtain a valid result about the research question that we would like to answer. This is because with two indicator variables, there are only four patterns (i.e., (0, 0), (0, 1), (1, 0), and (1, 1)) with four corresponding observed proportions. Similarly, for three indicators, there will be eight patterns. Therefore, there is only limited information and it is hard to satisfy the normality assumption of the underlying latent variables in the structural equation model. However, we did not find large scale simulation studies that address our concerns. The aim of the current paper is to fill this gap. Therefore, we conducted a simulation study to investigate the performance of SEM for the analysis of a small number of dichotomous indicator variables per factor.

In our simulation study, we conducted two types of experiments. In the first experiment, the performance of Confirmatory Factor Analysis (CFA) as a measurement model is studied. The outcome variables of interest for this experiment are the incidence of nonconvergence, occurrence of *Heywood* cases, and the quality of recovering the true factor scores in CFA. In the second experiment, we study the performance of the Multiple Indicators MultIple Causes (MIMIC) model (Stapleton, 1978; Kenneth, 1989; T. Brown, 2006). In this case, the outcome variables of interest are the type-I error rate and the power of the statistical test for the regression coefficients in the MIMIC model. In both experiments we study the impact of five design variables on the outcome variables. The design variables are type of indicator variables (i.e., interval or categorical), number of indicators, strength of factor structure, correlation between factor scores, and sample size.

The outline of this paper is as follows. In Section 2.2 we discuss issues with factor analysis and results found in the literature. The design and analysis of the simulation study is presented in Section 2.3. In Section 2.4, the results of the simulation studies are discussed. We conclude with a discussion of the results and some remarks for future research in Section 2.5.

2.2 Issues with Factor Models for Multivariate Data

2.2.1 Indeterminacy of Factor Scores

The indeterminacy of factor scores refers to a situation where the same indicator variables may produce different factor scores with the same model fit, and thus no unique solution does exist for the factor scores (Acito & Anderson, 1986; Guttman, 1955; Heermann, 1964, 1966; Schonemann, 1971; Schonemann & Wong, 1972; Green, 1976; Elffers, Bethlehem, & Gill, 1978). Some argue that the reason for the indeterminacy of factor scores is due to the presence of too many parameters compared to the number of equations in the model (Grice, 2001).

2.2.2 Improper Solutions

Factor analysis of multivariate data can sometimes produce improper solutions (Rindskopf, 1984; Boomsma, 1983, 1985; Kenneth, 1989; Chen, Bollen, Paxton, Curran, & Kirby, 2001). The most common improper solutions are nonconvergence and *Heywood* cases (Kenneth, 1989, pp. 282). *Heywood* cases occur when the estimated variances of a model become negative.

2.2.3 Previous Studies

Indeterminacy of factor scores in CFA has been studied by Acito and Anderson (1986). The impact of the number of indicators, factoring method, factor structure (i.e., the magnitude of factor loadings), number of factors, and sample size on indeterminacy of factor scores was investigated. Acito and Anderson found that both the factor structure and the factoring method have large effects on the indeterminacy of factor scores. A limitation of their study was that only interval indicators were considered. In the current study we also consider dichotomous indicators.

Improper solutions, i.e., nonconvergence and *Heywood* cases, in CFA has been studied using a Monte Carlo simulation by Anderson and Gerbing (1984) and by Boomsma (1985). Anderson and Gerbing studied the impact of sampling error and model characteristics on the incidence of improper solutions. Improper solutions occurred more frequently for smaller sample sizes and for models with fewer indicators for each factor (J. C. Anderson & Gerbing, 1984). Boomsma studied the impact of the number of indicators, correlation between factors, and factor structure. All of the design variables had a large effect on the incidence of improper solutions (Boomsma, 1985). Like the study by Acito and Anderson (1986), however, both studies considered only interval indicators. In this paper, we extend their study on improper solutions by including both interval and dichotomous indicator variables.

Marsh, Hau, Balla and Grayson (1998) performed a simulation study and studied extensively the impact of the number of indicators and sample size in a CFA on the occurrence of improper and nonconverged solutions, accuracy of parameter estimates, and goodness-of-fit indexes. Their main aim was to provide data driven evidence (contrary to rules of thumbs) for the number of indicators per factor and sample size to fit confirmatory factor models. They concluded that it is always good to have more indicators per factor and a larger sample size whenever possible. Similar studies were conducted by Ding, Velicer, and Harlow (1995), Kenny and McCaoch (2003) and Marsh, Balla, and McDonald (1988), although the main focus of these studies was on measures of fit for factor models.

In general, all these studies investigated the impact of variables of interest on statistical properties of CFA when the indicator variables are interval. Categorical (or dichotomous) indicator variables were not considered. Furthermore, emphasis was given for factor models and the MIMIC model was not studied in a similar fashion. Our present simulation study fills these gaps since both issues are addressed.

2.3 Monte Carlo Simulation Study

We followed the six-step approach of Monte Carlo simulation design in structural equation modelling (Paxton, Curran, Bollen, Kirby, & Fen, 2001; Skrondal, 2000; Boomsma, 2013).

2.3.1 The Research Problem

Our main objective is to investigate the performance of SEM models, specifically Confirmatory Factor Analysis (CFA) and the MIMIC model, with only a few dichotomous indicators assumed per factor. We investigate the quality of recovering the true factor scores and the incidence of improper solutions (nonconvergence and *Heywood* cases). We study the impact of five design variables on the outcome variables. To have a benchmark to compare the performance of latent variable models for analyzing dichotomous indicator variables, we also consider interval indicator variables. We used `Mplus` statistical software package (L. Muthen & Muthen, 1998-2012) with the default estimation procedures to analyze our simulated datasets, because that is how most applied researchers analyze their data.

The design variables are type of indicator variables (i.e., interval or categorical), number of indicators, strength of factor structure, strength of correlation between latent variables, and sample size. We conducted two experiments, where in the first experiment we study the performance of CFA and in the second experiment the performance of the MIMIC model.

2.3.2 Experimental Plan

The simulated data were generated from a 2-factor model whose path diagram is shown in Figure 2.1 for the CFA and in Figure 2.2 for the MIMIC model. In the path diagrams we have six observed variables y_j $(j = 1, 2, \ldots, 6)$, two latent variables θ_q $(q = 1, 2)$, and unique factors indicated by ϵ_j. The model parameters in CFA are the factor loadings λ_{jq} and the covariance between the latent variables ψ_{12}.

In the case of the MIMIC model (Figure 2.2) explanatory variables $(x_k, k = 1, 2)$ are added to the path diagrams. In addition to parameters of the factor model, the MIMIC model also has regression weights (γ_{kq}).

An equal number of indicator variables per factor was assumed in the data generation process. The variances of the factors were restricted to unity for identifiability of CFA. Table 2.2 shows the design variables considered in our Monte Carlo simulation and in the last column their corresponding values (or ranges) are given. The first design variable is type of indicators which is either dichotomous or interval. Two possible situations were considered for dichotomous indicator variables. The first case assumes a low success rate, i.e., between $5\% - 15\%$. This case is denoted by BLR in Table 2.2 which stands for Binary indicators with Low success Rates. The other situation has moderate success rate (between $40\% - 50\%$), and is denoted by BMR which stands for Binary indicators with Moderate success Rates.

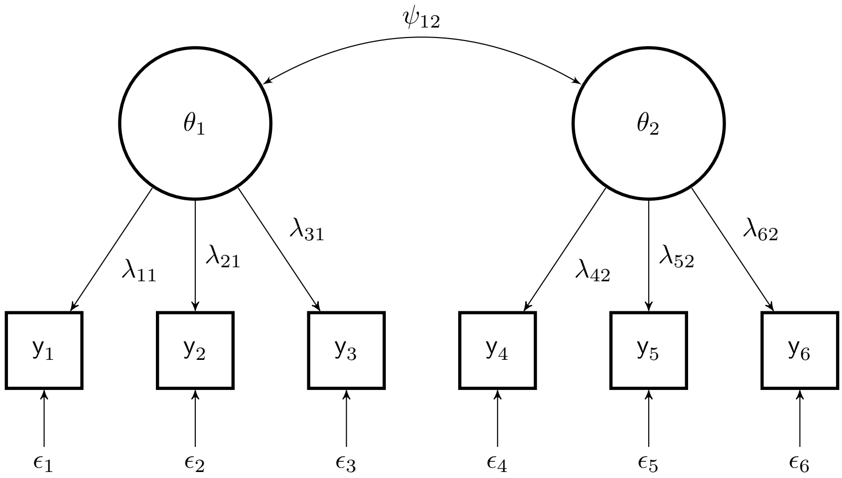

Figure 2.1: A path diagram of a factor model with six indicator variables represented by a square, and two latent variables represented by a circle.

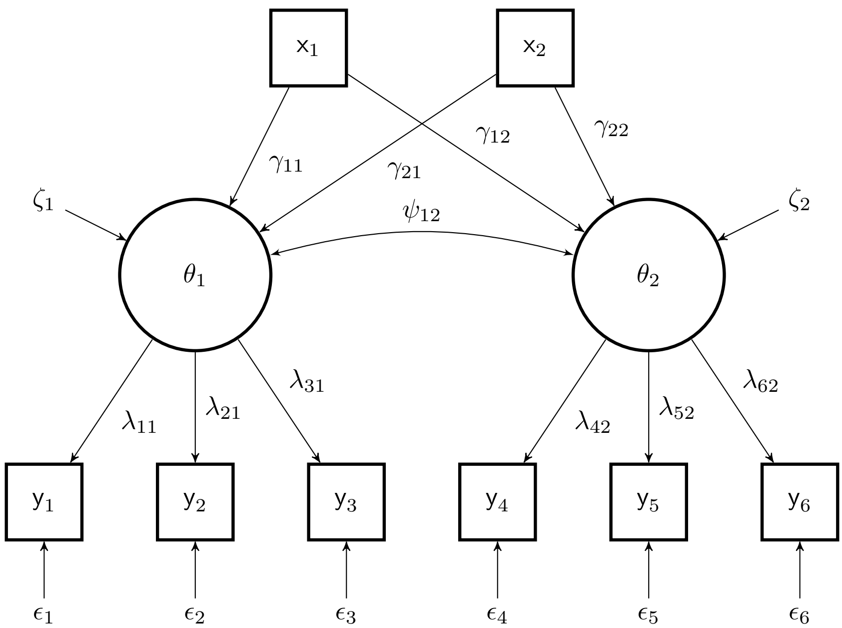

Figure 2.2: A path diagram for a MIMIC model with two external variables that are represented by a square.

Table 2.2: The design variables with their corresponding values (or ranges) that are considered in the Monte Carlo simulation study. BLR stands for Binary indicator variables with Low success Rates; and BMR for Binary indicator variables with Moderate success Rates.

Variable	Parameter	Level	Value/Range
Type of Indicators	–	BLR	$5\% - 15\%$
		BMR	$40\% - 50\%$
		Interval	–
Number of Indicators	J	Few	6
		Medium	10
		Large	16
Factor structure	λ_{jq}	Weak	$(0.316, 0.447)$
		Moderate	$(0.316, 0.632)$
		Strong	$(0.632, 0.775)$
Correlation between Factors	ψ_{12}	Independence	0.0
		Moderate	0.4
		Strong	0.8
Sample size	N	Very Small	50
		Small	100
		Big	300
		Very Big	$3,000$

For the number of indicators, Anderson and Gerbing (1984) suggested at least 3 indicators per factor in CFA. Kenny and McCoach (2003) varied the number of indicators from four to twenty-five to assess the impact on measures of fit. The number of indicators in our simulation study was varied from 3 to 8 per factor, which is equivalent to $J = 6$ to $J = 16$ indicators in total. For the factor structure, we used the ranges proposed by Acito and Anderson (1986). Both factor structure and variances for the measurement errors can be derived from the factor loadings, i.e., $\psi_{11} = \sum_{j=1}^{J} \lambda_{j1}^2$ and $\psi_{22} = \sum_{j=1}^{J} \lambda_{j2}^2$, and $\phi_j = 1 - \lambda_j^2$, respectively. In our simulation study, following Acito and Anderson (1986), we set factor loading values to: $(0.316, 0.447)$ for weak structure, $(0.316, 0.632)$ for moderate structure, and $(0.632, 0.775)$ for strong structure.

For the sample size, Boomsma (1985) recommended a sample size of at least $N = 50$ and Anderson and Gerbing (1984) suggested a sample size of at least $N = 150$. Boomsma and Hoogland (2001) showed that a sample size below $N = 200$ is vulnerable for the occurrence of improper solutions. In our Monte Carlo simulation the sample size was varied from $N = 50$ to $N = 3,000$. Three possible situations for the correlation between the latent variables were considered: $\psi_{12} = 0.0$ (independence), $\psi_{12} = 0.4$ (moderate association), and $\psi_{12} = 0.8$ (strong association).

2.3.3 Simulation

The simulated data is generated following the MIMIC model. In the simulated MIMIC model eight explanatory variables were considered. The true values that are used in the simulation study are based on the fitted MIMIC model on the NESDA data (Penninx et al., 2008). The first explanatory variable was generated from a Binomial distribution and the others from a Standard Normal distribution, i.e., $x_1 \sim \text{Bin}(0.67)$ and $x_k \sim N(0,1)$ for $k = 2, \ldots, 8$. The regression coefficients used in the simulation are the following, for x_1: $\gamma_{11} = \gamma_{12} = 0.00$; x_2: $\gamma_{21} = -0.10$, $\gamma_{22} = -0.20$; x_3: $\gamma_{31} = \gamma_{32} = 0.00$; x_4: $\gamma_{41} = 1.00$, $\gamma_{42} = 0.95$; x_5: $\gamma_{51} = -0.30$, $\gamma_{52} = -0.25$; x_6: $\gamma_{61} = \gamma_{62} = 0.00$; x_7: $\gamma_{71} = 0.00$, $\gamma_{72} = 0.10$; and x_8: $\gamma_{81} = \gamma_{82} = 0.00$.

Factor structures were then generated from a bivariate normal distribution $\boldsymbol{\theta} \sim \mathrm{N}_2(\boldsymbol{\mu}, \Psi)$, where

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1^{\mathsf{T}} \mathbf{x} + \zeta_1 \\ \boldsymbol{\gamma}_2^{\mathsf{T}} \mathbf{x} + \zeta_2 \end{bmatrix},$$

and

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1^{\mathsf{T}} \Sigma_x \boldsymbol{\gamma}_1 + \mathsf{Var}(\zeta_1) & \psi_{12} \\ \psi_{12} & \boldsymbol{\gamma}_2^{\mathsf{T}} \Sigma_x \boldsymbol{\gamma}_2 + \mathsf{Var}(\zeta_2) \end{bmatrix},$$

where $\boldsymbol{\gamma}_1$ is a vector of regression coefficients for the first factor, and similarly $\boldsymbol{\gamma}_2$ for the second factor.

2.3.4 Estimation

For each simulated data set a 2-factor model was fitted with and without explanatory variables, which corresponds to the CFA and the MIMIC model, respectively. The analysis was done using the `Mplus` statistical software version 7 (L. Muthen & Muthen, 1998-2012). A Maximum Likelihood Robust (MLR) estimator was employed for interval indicators whereas a Weighted Least Square estimator with Mean and Variance adjusted (WLSMV) was used for dichotomous indicator variables. The WLSMV is the default estimator in `Mplus`. We used the package called MplusAutomation (Hallquist, 2012) to help us call and run `Mplus` from the `R` environment.

The analysis procedure in our Monte Carlo simulation can be summarized as follows,

1. Fit a 2-factor CFA (or MIMIC model) on the simulated data.

2. Check if the fitted model is estimated without any problem due to improper solutions. Otherwise, identify the problem and record as nonconvergence and/or *Heywood*.

3. Estimate the factor scores, i.e., $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)$.

4. If the fitted model is estimated correctly, calculate the correlation between true and estimated factor scores, i.e., $\rho_q = \mathsf{Corr}(\theta_q, \hat{\theta}_q)$ where $q = 1, 2$. In the case of the MIMIC model, in addition to the correlation between factor scores, calculate:

(a) the type-I error rate the regression coefficients.

(b) the power of the regression coefficients.

5. Repeat Step 1 to 4 for each simulated data set.

2.3.5 Replication

In our Monte Carlo simulation we use a full factorial $3 \times 3 \times 3 \times 3 \times 4$ design, with in total 324 In each cell we use $R = 100$ replications.

2.3.6 Analysis of Output

For the first experiment, the variables of interest are the incidence of improper solutions and the quality of recovering the true factor scores in CFA .

The incidence of improper solutions was analyzed using a logistic regression model (Agresti, 2002). When a low rate of improper solutions is found in the data, Firth logistic regression (FLR: Firth, 1993) was used because it yields finite parameter estimates in the presence of complete or quasi-complete separation (Heinze & Schemper, 2002). SAS version 9.4 was used to fit the logistic regression models (SAS Institute Inc., 2013). The Odds Ratio (OR) was used as an effect size measure for evaluating the practical significance of the design variables and their interactions. We used the guidelines suggested by Ferguson (2009), i.e., an odds ratio of about 2 (or 0.50) indicates a small effect, about 3 (or 0.33) a medium effect, and about 4 (or 0.25) a large effect. For interpretation of simulation results, we focus on large effects for type of indicators and number of indicators, and their interactions with the other design variables.

Analysis of Variance (ANOVA) was used for analyzing the correlation data, i.e., ρ_1 and ρ_2, to assess the impact of design variables on the quality of recovering the true factor scores. Because ρ_1 and ρ_2 are very similar we focus on ρ_1. A Fisher's transformation is used to obtain an unbounded dependent variable, i.e., $z_1 = 0.5 \times \ln\left[(1+\rho_1)/(1-\rho_1)\right]$.

We used SPSS version 21 to fit the ANOVA model (IBM SPSS, 2012). The partial eta squared, denoted by η^2, will be used as a measure of effect size for the ANOVA model. According to Cohen (1988), a value of $\eta^2 = 0.01$ indicates a small effect, $\eta^2 = 0.059$ a medium effect, and $\eta^2 = 0.138$ a large effect. For interpretation of simulation results, we focus on large effects for type of indicators and number of indicators, and their interactions with the other design variables.

In the second experiment we are further interested in the type-I error rate and the power for the regression weights of the MIMIC model. These measures were obtained by first calculating the proportion of cases in which an effect becomes statistically significant. For the effects equal to zero, the calculated proportion represents the type-I error rate; otherwise, the proportion corresponds to the power of the test. In the case of type-I error rate, a 95% confidence interval of the proportion using the Wilson interval was calculated (L. D. Brown, Cai, & DasGupta, 2001).

2.4 Results

2.4.1 Experiment-I: Confirmatory Factor Analysis

Nonconvergence in CFA

About 18.9% of the analyses in our simulation study did not converge. We applied logistic regression on the nonconvergence outcome variable (1: not converged; 0: converged) to investigate the impact of the design variables. The observed proportions of nonconvergence cross classified by design variables are presented in Table 2.3. A two-way interaction logistic model was fitted to the nonconvergence data.

The results of the 2-way interaction model are displayed in the Appendix (Table A.1); our focus here will be on the effects of two of the design variables, i.e., type of indicators and number of indicators, and their interaction with the other design variables. The type

Table 2.3: Percentage of nonconvergence in CFA under different experimental settings. Each cell result is based on $R = 100$ simulated replications.

			Sample Size											
			50			100			300			3000		
			Number of Indicators											
Type of Indicators	Factor structure	Correlation between factors	6	10	16	6	10	16	6	10	16	6	10	16
BLR	Weak	Independence	73.0	52.0	54.0	81.0	69.0	43.0	79.0	58.0	7.0	13.0	0.0	0.0
		Moderate	78.0	70.0	51.0	81.0	71.0	41	76.0	33.0	10.0	2.0	0.0	0.0
		Strong	72.0	73.0	76.0	87.0	65.0	48.0	64.0	36.0	23.0	15.0	2.0	0.0
	Moderate	Independence	79.0	58.0	50.0	83.0	65.0	31.0	75.0	24.0	5.0	5.0	0.0	0.0
		Moderate	76.0	63.0	42.0	82.0	48.0	21.0	56.0	10.0	1.0	0.0	0.0	0.0
		Strong	79.0	63.0	62.0	81.0	56.0	35.0	45.0	21.0	12.0	6.0	1.0	0.0
	Strong	Independence	67.0	56.0	36.0	66.0	26.0	9.0	27.0	1.0	0.0	0.0	0.0	0.0
		Moderate	74.0	58.0	41.0	69.0	33.0	21.0	12.0	1.0	0.0	0.0	0.0	0.0
		Strong	71.0	62.0	42.0	54.0	38.0	25.0	7.0	1.0	0.0	0.0	0.0	0.0
BMR	Weak	Independence	79.0	51.0	27.0	75.0	41.0	7.0	48.0	8.0	0.0	0.0	0.0	0.0
		Moderate	85.0	60.0	31.0	68.0	47.0	6.0	31.0	2.0	0.0	0.0	0.0	0.0
		Strong	79.0	56.0	46.0	66.0	33.0	27.0	29.0	17.0	6.0	4.0	0.0	0.0
	Moderate	Independence	82.0	49.0	15.0	73.0	23.0	2.0	32.0	0.0	0.0	0.0	0.0	0.0
		Moderate	75.0	30.0	12.0	56.0	7.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0
		Strong	66.0	52.0	24.0	50.0	20.0	8.0	18.0	3.0	0.0	0.0	0.0	0.0
	Strong	Independence	52.0	22.0	7.0	22.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
		Moderate	52.0	14.0	2.0	26.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
		Strong	28.0	9.0	5.0	10.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Interval	Weak	Independence	36.0	25.0	7.0	16.0	3.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0
		Moderate	27.0	22.0	5.0	11.0	4.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0
		Strong	23.0	15.0	2.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Moderate	Independence	32.0	9.0	1.0	9.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
		Moderate	12.0	4.0	0.0	7.0	1.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0
		Strong	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Strong	Independence	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Moderate	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Strong	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

of indicators has a large effect on the incidence of nonconvergence in CFA. Moreover, we found a large effect of 2-way interaction between the type of indicators and the following variables: number of indicators, factor structure, and sample size. There is also a large effect of number of indicators on the incidence of nonconvergence in CFA, and its 2-way interaction with the sample size. Figure 2.3 displays the corresponding interaction plots for the large effects. The first three panels (from left to right) show interaction plot between the type of indicators and the other design variables (i.e., the number of indicators, the factor structure, and the sample size). The last panel is for the interaction plot between the number of indicators and the sample size.

Regardless of the other design variables (i.e., number of indicators, factor structure, and sample size), we found a large effect of type of indicators on the prevalence of nonconvergence in CFA. The worst result was obtained for the binary indicators, specifically for the BLR data. For interval indicators, there was not much effect of the other design variables on the prevalence of nonconvergence in CFA.

By looking at the first and the last panel in Figure 2.3, there is a large interaction effect between the number of indicators with the type of indicators and the sample size. That is, the worst prevalence of nonconvergence due to the number of indicators was obtained when the binary indicators (i.e., BLR and BMR data) are analyzed by CFA. For the sample size, the worst prevalence of nonconvergence due to the number of indicators was found when the sample size is below $N <= 300$.

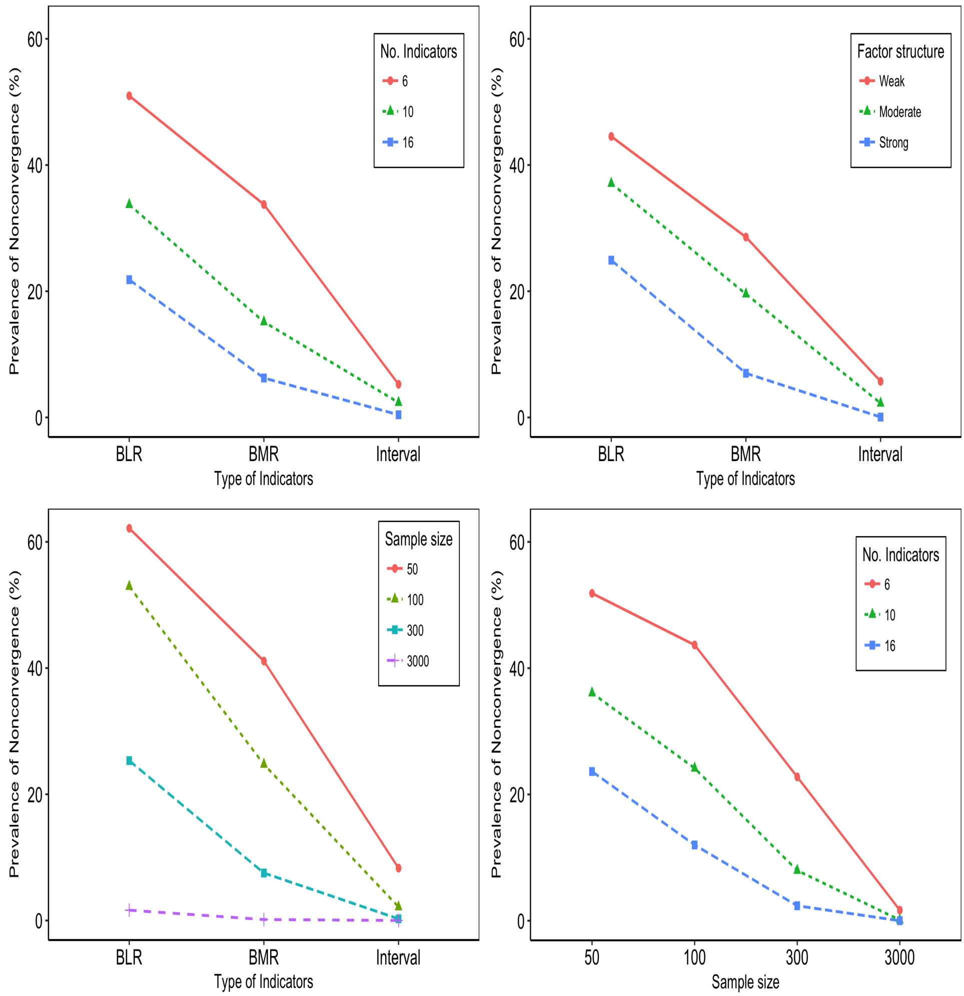

Figure 2.3: Interaction plot for Nonconvergence rate: The first three panels (from left to right) show interaction plot between the type of indicators and the number of indicators, the factor structure, and the sample size, respectively. The last panel is for the interaction between the number of indicators and the sample size.

Heywood cases in CFA

About 6.6% of the analyses in our simulation study resulted in *Heywood* cases. We applied logistic regression models on the *Heywood* outcome variable (1: yes; 0: no). The observed proportions of *Heywood* cases cross classified by the design variables are shown in Table 2.4. A 2-way interaction logistic model was fitted on the *Heywood* data, and the results are presented in the Appendix (Table A.2).

Like for the nonconvergence analysis, our focus will be on the effects of the type of indicators and the number of indicators, and their interaction with the other design variables. The type of indicators has a large effect on the incidence of *Heywood* in CFA. Moreover, we found a large effect of 2-way interaction between the type of indicators and all the other design variables, i.e., number of indicators, factor structure, correlation between latent variables, and sample size. There is also a large effect of number of indicators on the incidence of nonconvergence in CFA, and its 2-way interaction with both the factor structure and sample size. Figure 2.4 displays the interaction plots for the large effects. The first four panels (from left to right) show interaction plot between the type of indicators with the number of indicators, the factor structure, the correlation between underlying latents, and the sample sizes. The last two panels are for the interaction plot between the number of indicators with the factor structure and sample size.

In the first three panels, it can be seen that there is no large difference in prevalence of *Heywood* cases among the type of indicators used in CFA. The fourth panel shows the interaction with the sample size, where the highest number of *Heywood* cases was found for the BLR data, except for a large and small data sets.

There is a large effect of the number of indicators on the prevalence of *Heywood* cases in CFA. The first panel in the last row shows that the worst result was found for a small number of indicators regardless of the type of indicators in CFA. Furthermore, more *Heywood* cases were found for the smallest number of indicators with weak factor structure and with small sample sizes.

Table 2.4: Percentage of *Heywood* cases in CFA under different experimental settings. Each cell result is based on $R = 100$ simulated replications.

			Sample Size											
			50			100			300			3000		
			Number of Indicators											
Type of Indicators	Factor structure	Correlation between factors	6	10	16	6	10	16	6	10	16	6	10	16
BLR	Weak	Independence	21.0	6.0	3.0	51.0	24.0	2.0	50.0	26.0	3.0	5.0	0.0	0.0
		Moderate	21.0	3.0	6.0	41.0	23.0	7.0	44.0	15.0	4.0	1.0	0.0	0.0
		Strong	14.0	4.0	2.0	42.0	9.0	1.0	28.0	6.0	1.0	0.0	0.0	0.0
	Moderate	Independence	18.0	2.0	4.0	52.0	16.0	0.0	48.0	9.0	0.0	0.0	0.0	0.0
		Moderate	9.0	3.0	1.0	33.0	7.0	0.0	26.0	1.0	0.0	0.0	0.0	0.0
		Strong	8.0	0.0	4.0	31.0	4.0	0.0	8.0	0.0	0.0	0.0	0.0	0.0
	Strong	Independence	10.0	4.0	7.0	12.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0
		Moderate	6.0	3.0	4.0	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Strong	5.0	2.0	2.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BMR	Weak	Independence	47.0	13.0	3.0	53.0	18.0	3.0	34.0	5.0	0.0	0.0	0.0	0.0
		Moderate	45.0	20.0	2.0	37.0	13.0	1.0	12.0	0.0	0.0	0.0	0.0	0.0
		Strong	51.0	14.0	1.0	25.0	2.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
	Moderate	Independence	41.0	10.0	1.0	38.0	6.0	0.0	11.0	0.0	0.0	0.0	0.0	0.0
		Moderate	38.0	5.0	0.0	23.0	1.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0
		Strong	17.0	1.0	0.0	13.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
	Strong	Independence	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Moderate	3.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Strong	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Interval	Weak	Independence	66.0	48.0	14.0	54.0	14.0	2.0	20.0	0.0	0.0	0.0	0.0	0.0
		Moderate	51.0	34.0	12.0	46.0	7.0	1.0	9.0	0.0	0.0	0.0	0.0	0.0
		Strong	38.0	18.0	6.0	13.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Moderate	Independence	60.0	22.0	2.0	40.0	2.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0
		Moderate	40.0	13.0	1.0	24.0	1.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0
		Strong	14.0	4.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Strong	Independence	8.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Moderate	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Strong	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

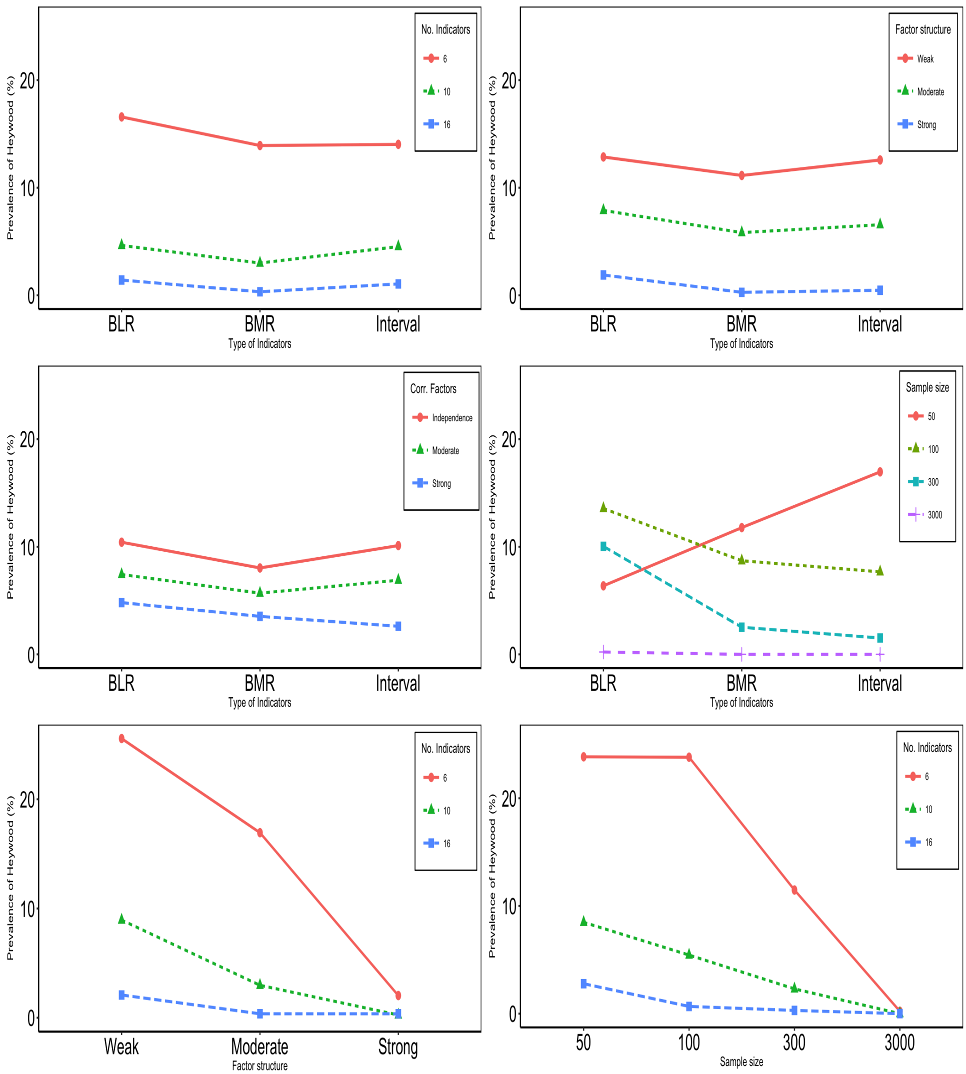

Figure 2.4: Interaction plot for Heywood rate: The first four panels (from left to right) show the interaction between the type of indicators and the number of indicators, the factor structure, the correlation between underlying latents, and the sample size, respectively. The last two panels are for the interaction between the number of indicators and the factor structure and the sample size.

Recovery of Factor Scores

The observed average correlations between the true and estimated factor scores for the first latent variable cross classified by the design variables as displayed in Table 2.5. A 2-way interaction ANOVA model was fitted on the transformed correlation data, and the results are presented in the Appendix (Table A.3). Large effects were found for the number of indicators, the type of indicators, and the interaction between type of indicators and factor structure. Figure 2.5 displays the interaction plots for the large effects. The first panel shows two main-effect plots corresponding to effects of the type of indicators and the number of indicators, respectively. The second panel shows the interaction plot between the type of indicators and the factor structure.

Both plots show that there is a large effect of type of indicators on the quality of recovering the true factor scores in CFA. The worst result was obtained for the binary indicators (i.e., BLR and BMR) regardless of the other design variables.

The first plot also shows a large effect of the number of indicators on the quality of recovering the factor scores in CFA. The worst result was found for the smallest number of indicators. Although we found a large interaction effect between the type of indicators and the factor structure, the second plot in Figure 2.5 does not clearly show this effect.

Table 2.5: Quality of Recovering the True Factor scores: Average correlation between the true and estimated factor scores of CFA, i.e., Corr$(\theta_1, \hat{\theta}_1) = \hat{\rho}_1$, under different experimental settings. Each cell represents the results of $R = 100$ simulated replications, except those models that were not identified due to improper solutions.

			Sample Size											
			50			100			300			3000		
			Number of Indicators											
Type of Indicators	Factor structure	Correlation between factors	6	10	16	6	10	16	6	10	16	6	10	16
BLR	Weak	Independence	0.29	0.32	0.37	0.29	0.32	0.40	0.23	0.38	0.47	0.35	0.44	0.53
		Moderate	0.30	0.42	0.49	0.29	0.37	0.46	0.32	0.41	0.50	0.36	0.47	0.55
		Strong	0.39	0.47	0.56	0.36	0.41	0.54	0.38	0.48	0.58	0.44	0.53	0.62
	Moderate	Independence	0.32	0.38	0.51	0.39	0.46	0.53	0.38	0.51	0.60	0.44	0.53	0.62
		Moderate	0.42	0.50	0.57	0.41	0.50	0.59	0.43	0.53	0.62	0.46	0.55	0.64
		Strong	0.46	0.56	0.65	0.52	0.56	0.65	0.49	0.59	0.68	0.52	0.62	0.69
	Strong	Independence	0.48	0.61	0.70	0.54	0.65	0.72	0.57	0.66	0.73	0.59	0.66	0.73
		Moderate	0.57	0.64	0.72	0.59	0.67	0.73	0.60	0.67	0.74	0.61	0.69	0.75
		Strong	0.61	0.70	0.76	0.64	0.69	0.77	0.65	0.72	0.78	0.66	0.73	0.78
BMR	Weak	Independence	0.36	0.46	0.50	0.43	0.50	0.57	0.43	0.54	0.65	0.47	0.58	0.67
		Moderate	0.44	0.42	0.56	0.43	0.52	0.63	0.47	0.58	0.67	0.50	0.60	0.68
		Strong	0.49	0.51	0.67	0.52	0.60	0.70	0.53	0.64	0.73	0.57	0.66	0.74
	Moderate	Independence	0.47	0.58	0.68	0.49	0.63	0.74	0.56	0.66	0.75	0.58	0.68	0.76
		Moderate	0.53	0.62	0.70	0.58	0.65	0.75	0.57	0.68	0.76	0.60	0.69	0.77
		Strong	0.56	0.71	0.77	0.63	0.70	0.79	0.65	0.74	0.81	0.66	0.75	0.81
	Strong	Independence	0.73	0.80	0.87	0.74	0.82	0.87	0.75	0.82	0.87	0.75	0.83	0.88
		Moderate	0.74	0.81	0.87	0.75	0.82	0.88	0.76	0.83	0.88	0.77	0.83	0.88
		Strong	0.79	0.85	0.89	0.79	0.85	0.90	0.80	0.86	0.90	0.81	0.86	0.90
Interval	Weak	Independence	0.41	0.54	0.66	0.46	0.64	0.73	0.54	0.67	0.75	0.58	0.68	0.76
		Moderate	0.53	0.59	0.68	0.53	0.64	0.74	0.57	0.68	0.76	0.60	0.70	0.77
		Strong	0.52	0.65	0.77	0.61	0.71	0.79	0.64	0.73	0.81	0.66	0.75	0.81
	Moderate	Independence	0.60	0.73	0.82	0.65	0.75	0.83	0.66	0.78	0.84	0.69	0.78	0.85
		Moderate	0.66	0.75	0.82	0.67	0.76	0.85	0.69	0.78	0.84	0.71	0.79	0.85
		Strong	0.69	0.80	0.86	0.73	0.81	0.87	0.75	0.82	0.88	0.76	0.83	0.88
	Strong	Independence	0.84	0.91	0.94	0.85	0.90	0.94	0.86	0.91	0.94	0.87	0.91	0.94
		Moderate	0.84	0.90	0.94	0.86	0.91	0.94	0.87	0.91	0.94	0.87	0.92	0.94
		Strong	0.87	0.92	0.95	0.88	0.92	0.95	0.89	0.93	0.95	0.89	0.93	0.95

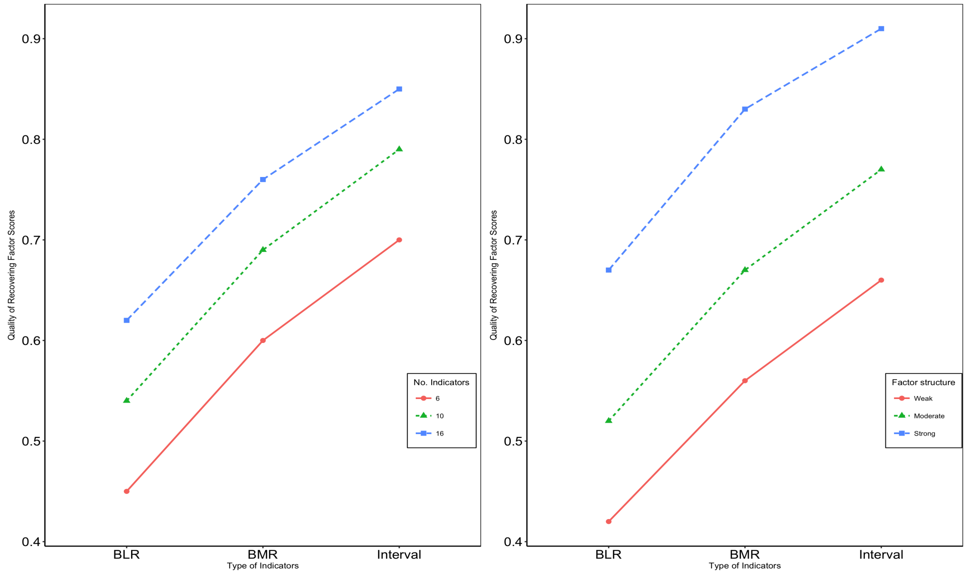

Figure 2.5: Interaction plot for Quality of Recovering Factors: The first panel shows two main effects for the type and number of indicators. The second panel shows the interaction between the type of indicators and the factor structure.

2.4.2 Experiment-II: The MIMIC Model

In the second experiment, we study the type-I error rate and the power of the regression parameters (i.e., γ) of the MIMIC model.

Type-I Error

In this section we study the impact of design variables on the Type-I error rate. Although there are a couple of parameters whose true values are set to zero in the simulation study, for demonstration purpose we chose one of the parameters, i.e., γ_{31} which indicates the relationship between X_3 and the first latent variable whose results are presented in Table 2.6.

Those values whose 95% confidence interval excluded the nominal level of significance ($\alpha = 0.05$) were made bold; there are a total of fourteen cells which resulted in such cases,

Table 2.6: Observed type-I error rates for the relationship between X_3 and the first factor, γ_{31}. The values in bold represent 95% confidence interval excluding the nominal level of significance ($\alpha = 0.05$). The number of replications per cell differs because of improper solutions. Dashed lines indicate no valid results were obtained for that cell.

			Sample Size											
			50			100			300			3000		
			Number of Indicators											
Type of Indicators	Factor structure	Correlation between factors	6	10	16	6	10	16	6	10	16	6	10	16
BLR	Weak	Independence	–	–	–	0.00	0.00	0.00	0.01	0.04	0.02	0.01	0.01	0.04
		Moderate	–	–	–	0.00	0.03	0.00	0.02	0.02	0.07	0.03	0.02	0.03
		Strong	–	–	–	0.00	0.00	0.03	0.01	0.05	0.02	0.05	0.05	0.04
	Moderate	Independence	–	–	–	0.00	0.00	0.00	0.00	0.09	0.06	**0.10**	0.03	0.03
		Moderate	–	–	–	0.00	0.00	0.00	0.04	0.04	0.01	0.03	0.03	0.02
		Strong	0.00	–	–	0.00	0.00	0.17	0.01	0.01	0.03	0.08	0.06	0.02
	Strong	Independence	–	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.10	0.09	0.09	0.00
		Moderate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.04	0.08	0.07	0.00
		Strong	–	0.00	–	0.00	0.00	0.00	0.00	0.00	0.11	0.03	0.08	0.02
BMR	Weak	Independence	0.07	0.02	0.02	0.01	0.06	0.02	0.04	0.08	0.02	0.05	0.01	0.04
		Moderate	0.03	0.00	0.07	0.02	0.05	0.03	0.02	0.05	0.05	0.08	0.06	0.05
		Strong	0.00	0.02	0.07	0.06	0.05	0.06	0.03	0.06	0.03	0.09	0.05	0.03
	Moderate	Independence	0.05	0.10	0.00	0.03	0.06	0.03	0.03	0.05	0.04	0.08	0.02	0.08
		Moderate	0.00	0.03	0.08	0.05	0.03	0.05	0.03	0.04	0.11	0.05	0.02	0.02
		Strong	0.03	0.04	0.05	0.04	0.04	0.07	0.02	0.05	0.05	0.08	0.03	0.03
	Strong	Independence	0.00	0.00	0.00	0.00	0.00	0.07	0.05	0.00	0.00	0.00	0.00	0.00
		Moderate	0.00	0.00	–	0.00	0.00	0.11	0.00	0.00	0.00	0.05	0.06	0.00
		Strong	0.00	–	0.00	0.20	0.00	0.00	0.07	**0.38**	0.10	0.03	0.06	0.04
Interval	Weak	Independence	0.03	0.07	0.09	0.01	0.08	0.05	0.08	0.07	0.04	0.08	0.05	0.03
		Moderate	0.07	**0.11**	**0.11**	0.07	0.08	0.00	0.02	0.04	0.04	0.07	0.07	0.06
		Strong	0.04	0.05	0.06	0.09	0.07	**0.11**	0.03	0.04	0.05	0.03	0.06	0.05
	Moderate	Independence	0.08	0.09	**0.12**	**0.11**	**0.10**	0.05	0.05	0.06	0.05	0.04	0.06	0.01
		Moderate	**0.10**	0.09	0.09	**0.10**	0.05	0.05	0.01	0.06	0.07	0.03	0.04	0.04
		Strong	0.04	**0.10**	0.08	0.03	0.04	0.05	0.06	0.03	0.03	0.02	0.05	0.08
	Strong	Independence	**0.10**	0.09	0.07	0.08	0.06	0.08	0.03	0.06	**0.11**	0.02	0.02	0.02
		Moderate	0.08	0.09	0.07	0.04	0.06	0.05	0.03	0.06	0.05	0.06	0.04	0.04
		Strong	0.09	**0.14**	0.07	0.07	0.08	0.09	0.07	0.08	0.06	0.06	0.04	0.04

i.e., one for BLR data, one for BMR data, and the rest for interval data. However, for BLR and BMR data with $N = 50$ no results were found due to the presence of improper solutions and are thus represented by dashed lines in Table 2.6. Therefore, this resulted in a wider confidence interval for the binary indicators (i.e., the BLR and the BMR data) and because of that the 5% level of significance level is included.

Although the confidence intervals obtained for both BLR and BMR data do include the level of significance, most of the point estimates are either zero or very high. Therefore, the observed type-I error rates were too conservative (i.e., values very close to zero) for BLR, and inflated (i.e., above the 5% nominal level of significance) for BMR indicators. In the case of interval indicators, most of the confidence intervals included the nominal level of significance, and their results seems stable around $\alpha = 0.05$.

The number of indicators had an impact on the recovery of type-I error rate, particularly for dichotomous indicator variables with low success rates (BLR). That is, among the 108 observed type-I error rates obtained from BLR only thirteen included the 5% nominal level of significance in their 95% confidence interval for $J = 6$, while sixteen for $J = 10$ and nineteen for $J = 16$ recovered the required type-I error rate. In the case of both BMR and interval indicators, we found no significant difference on the recovery of type-I error rate among the type of indicators.

Power

In this section we focus on the observed power of three of the parameters in the MIMIC model, i.e., $\gamma_{51} = -0.30$ representing a moderate effect, $\gamma_{72} = 0.10$ for a small effect, and $\gamma_{42} = 0.95$ for a strong effect. Their results are shown in Table 2.7 and in the Appendix (Table A.4 and A.5), respectively. Some of the cells of these tables had no values due to the presence of improper solutions and are consequently represented by dashed lines. A threshold value of 0.80 will be used as a criterion to represent adequate power.

It is evident from Table 2.7 that most of the observed power values obtained from

Table 2.7: Observed power for the relationship between X_5 and the first factor, $\gamma_{51} = -0.30$. The number of replications per cell differ because of improper solutions. Dashed lines indicate no valid results were obtained for that cell.

			Sample Size											
			50			100			300			3000		
			Number of Indicators											
Type of Indicators	Factor structure	Correlation between factors	6	10	16	6	10	16	6	10	16	6	10	16
BLR	Weak	Independence	–	–	–	0.12	0.06	0.03	0.27	0.31	0.43	0.99	1.00	1.00
		Moderate	–	–	–	0.00	0.00	0.00	0.29	0.34	0.56	1.00	1.00	1.00
		Strong	–	–	–	0.09	0.03	0.03	0.22	0.35	0.42	1.00	1.00	1.00
	Moderate	Independence	–	–	–	0.00	0.00	0.00	0.21	0.34	0.43	1.00	1.00	1.00
		Moderate	–	–	–	0.07	0.00	0.00	0.17	0.41	0.38	1.00	1.00	1.00
		Strong	0.00	–	–	0.05	0.06	0.08	0.30	0.38	0.42	1.00	1.00	1.00
	Strong	Independence	–	0.00	0.00	0.00	0.00	0.00	0.06	0.47	0.51	1.00	1.00	1.00
		Moderate	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.53	0.38	1.00	1.00	1.00
		Strong	–	0.00	–	0.25	0.00	0.00	0.20	0.40	0.26	1.00	1.00	1.00
BMR	Weak	Independence	0.10	0.08	0.11	0.11	0.23	0.31	0.53	0.64	0.69	1.00	1.00	1.00
		Moderate	0.06	0.13	0.07	0.21	0.29	0.23	0.52	0.65	0.71	1.00	1.00	1.00
		Strong	0.06	0.11	0.12	0.16	0.26	0.31	0.60	0.59	0.65	1.00	1.00	1.00
	Moderate	Independence	0.09	0.03	0.00	0.26	0.21	0.26	0.53	0.73	0.72	1.00	1.00	1.00
		Moderate	0.16	0.08	0.25	0.23	0.33	0.37	0.52	0.68	0.70	1.00	1.00	1.00
		Strong	0.15	0.11	0.08	0.21	0.30	0.18	0.53	0.67	0.72	1.00	1.00	1.00
	Strong	Independence	0.00	0.00	0.33	0.14	0.50	0.13	0.67	0.80	0.35	1.00	1.00	1.00
		Moderate	0.00	0.00	–	0.18	0.33	0.11	0.64	0.71	0.73	1.00	1.00	1.00
		Strong	0.00	–	0.00	0.60	0.50	0.50	0.71	0.50	0.70	1.00	1.00	1.00
Interval	Weak	Independence	0.12	0.21	0.22	0.21	0.28	0.23	0.58	0.69	0.79	1.00	1.00	1.00
		Moderate	0.17	0.23	0.15	0.26	0.27	0.25	0.69	0.72	0.85	1.00	1.00	1.00
		Strong	0.16	0.24	0.23	0.21	0.26	0.31	0.59	0.67	0.71	1.00	1.00	1.00
	Moderate	Independence	0.13	0.28	0.21	0.25	0.26	0.38	0.65	0.75	0.86	1.00	1.00	1.00
		Moderate	0.21	0.22	0.17	0.34	0.43	0.44	0.64	0.73	0.82	1.00	1.00	1.00
		Strong	0.22	0.24	0.27	0.32	0.42	0.32	0.75	0.72	0.86	1.00	1.00	1.00
	Strong	Independence	0.19	0.34	0.26	0.35	0.35	0.42	0.81	0.84	0.84	1.00	1.00	1.00
		Moderate	0.30	0.25	0.26	0.33	0.35	0.38	0.80	0.86	0.78	1.00	1.00	1.00
		Strong	0.17	0.29	0.34	0.39	0.30	0.48	0.86	0.77	0.83	1.00	1.00	1.00

dichotomous indicators are very low, except when the sample size is very large. Let us elaborate on this point using the results under a sample size of $N = 300$ from Table 2.7. Out of the twenty-seven observed values for BLR (i.e., Binary indicators with Low success Rates) none of them passed the threshold value of 0.80 and the maximum value achieved was only 0.56. For the case of BMR (i.e., Binary indicators with Moderate success Rates), only one out of twenty-seven had an observed power of exactly 0.80. In the case of interval indicators, however, eleven out of twenty-seven satisfied the criteria and the maximum value achieved was an observed power of 0.86.

The minimum power was 0.00 for both BLR and BMR while it was 0.12 for interval indicators. With a small effect size of $\gamma_{72} = 0.10$ as shown in the Appendix (Table A.4), among the 216 cells for dichotomous indicators (i.e., both BLR and BMR) only three of the observed values are larger than 0.8 while in the case of interval indicators nineteen out of 108 cells satisfied this criterion. These results show that the MIMIC model performed poorly for analyzing dichotomous indicators, particularly for dichotomous indicators with low success rates.

2.5 Conclusion and Discussion

Structural equation models are originally proposed for analysis of continuous (or interval) indicator variables. Recently, factor analysis and structural equation models have been applied for data with dichotomous indicators and with only a few indicators per latent variable, i.e., 2 or 3 (Krueger, 1999; Beesdo-Baum et al., 2009). Using a Monte Carlo simulation study, we showed that latent variable models applied on such type of data performed poorly with higher incidence of improper solutions, poor quality of recovering the true factor scores, too conservative or inflated type-I error rates, and weak power.

About 18.9% out of all the analyses in the CFA did not achieve convergence, and about 6.6% were *Heywood* cases (i.e., out-of-bound problem). It is shown that the type of indicators and the number of indicators in CFA plays a major role on the occurrence of the nonconvergence in CFA. That is, high prevalence of nonconvergence was obtained for binary indicators with a few indicators per latent variable. For the occurrence of *Heywood* cases in CFA, the number of indicators also played a major role. The quality of recovering the true factor scores in CFA was poor in the case of binary indicators, and it became worse with less indicators per latent variable.

We evaluated the performance of the MIMIC model using the type-I error rate and the power of test for the regression weights. Most of the confidence intervals of the type-I error distribution obtained from the dichotomous indicators, did not include the 5% nominal level of significance. The type-I error rates were mostly conservative, although few of them were inflated. For interval indicators, however, most of the results included the nominal level of significance within their 95% confidence interval. The power of the test with dichotomous indicators was poor compared to the interval indicators.

It is important to note that we used an advantageous design for our Monte Carlo simulation study. The latent variables were generated from a bivariate normal distribution. Moreover, the population model was correctly specified. In empirical studies it is likely that assumptions are only partially valid. Moreover, the fitted model could be misspecified; for example, an important indicator variable may not have been included in the analysis. Under such conditions we would expect even more improper solutions and factor scores that are further off than what we found in our current study.

Latent trait or Item Response Theory (IRT) model has also been proposed for analyzing dichotomous indicator variables (Lord & Novick, 1968). It was shown by Mislevy (1986) and Takane and De Leeuw (1987), that CFA and IRT models are formally equivalent and thus yielding similar results. Knol and Berger (1991) conducted a simulation study to compare the performance of these models and found that the common factor analysis on the matrix of tetrachoric correlations performed similar to IRT models for multidimensional data. Furthermore, Glöckner-Rist and Hoijtink (2003) recommended a joint application of the models. We conclude our discussion with a recommendation for those researchers who do confirmatory factor analysis on data with a small number of dichotomous indicator variables. It is shown in our Monte Carlo simulation that the method performed poorly for this type of data and therefore must be used carefully. An alternative statistical model which requires less assumptions might be more appropriate, for example the multivariate logistic distance model (Worku & De Rooij, 2018).