

UNIVERSITY OF JYVÄSKYLÄ  
DEPARTMENT OF MATHEMATICS  
AND STATISTICS

REPORT 111

UNIVERSITÄT JYVÄSKYLÄ  
INSTITUT FÜR MATHEMATIK  
UND STATISTIK

BERICHT 111

# LATENT GROWTH MIXTURE MODELING: A SIMULATION STUDY

ASKO TOLVANEN



JYVÄSKYLÄ  
2007

Editor: Pekka Koskela  
Department of Mathematics and Statistics  
P.O. Box 35 (MaD)  
FI-40014 University of Jyväskylä  
Finland

ISBN 978-951-39-2971-8  
ISSN 1457-8905

Copyright © 2007, by Asko Tolvanen  
and University of Jyväskylä

University Printing House  
Jyväskylä 2007

## ABSTRACT

Latent growth curve modeling (LGM) combined with the latent classes (LGMM) in the SEM context, is the method under investigation in this study. This dynamic way of analyzing longitudinal data takes an increasingly central position in the social sciences, e.g. in psychology. Despite twenty years development of the theory behind the LGM and LGMM, these are novel methods in analyzing data in practice. With limited sample size the functionality of the model is unknown.

The aim of this dissertation was to examine the functionality of the linear LGM model with four repeated measurements, which is a typical case in longitudinal research. LGMM parameters were estimated using maximum likelihood estimation with robust standard errors (MLR). The effect of differences between latent classes in mean values of latent components with varying sample sizes is examined in this study. Other affecting factors examined are reliability of observed variables, number of repeated measures, model construct and additional measurement points. The functionality of LGMM was approached from three different viewpoints: 1) problems in estimation of model parameters expressed as number of failed estimations and as the number of negative variance estimates, 2) the ability of AIC, BIC and aBIC information criteria and VLMR, LMR and BLRT statistical tests to decide the number of latent classes, and 3) good parameter estimation, which was evaluated using four different criteria: MSE, proportion of bias in MSE, bias of standard error, and 95 % coverage.

The results of Monte Carlo simulations suggest that from information criteria AIC, BIC aBIC and VLMR and LMR tests, BIC is most useful with small sample sizes ( $n < 500$ ) and aBIC with large sample sizes ( $n \geq 500$ ). The few results suggest that the BLRT test could be useful in any situation. More investigation is needed to further support the functionality of this test. The study reveals that the estimation of LGMM fails only in a few cases, and problems in estimation appear mainly in the negative variance estimates. The results of the simulations suggest that it is possible to identify the true two-latent classes when SMD is at least 2, in which case reliability of observed variables should be high and the sample size should be relatively large. In that case estimation produce good parameter estimates. When SMD is 4 or 5, the probability in identifying the right two-latent-class solution instead of the wrong one-class solution is greater than .70 with the smallest sample size ( $n=50$ ) using BIC in models with high reliability. To achieve reliable results in estimation, the sample size should be greater than 50.

Key words: Latent growth mixture modeling, Monte Carlo simulation

## ACKNOWLEDGEMENTS

I am highly impressed about the way that Professor Esko Leskinen has supervised me through this study. His endless enthusiasm to advise and solve theoretical as well as practical problems that emerged during the process has encouraged me to bring this research into completion. I am deeply indebted to him for his help.

I want to thank Professor Markku Rahiala and Docent Erkki Alanen for reviewing my thesis. I am very grateful to Docent Kaisa Aunola for her invaluable contributions to the editing of the work. Docent Kaisa Aunola and Professor Jari-Erik Nurmi read the preliminary text and gave constructive comments to the language and structure of the thesis. I extend my warm thanks to Katriina Hyvönen and Kiran Kamat for revising the language of this thesis.

I wish to thank all my colleagues at the Department of Psychology and the Psykocenter for providing a creative and inspiring working environment. I also wish to express my gratitude to Professor Lea Pulkkinen, Professor Jari Nurmi and Professor Heikki Lyytinen and leader of the Department of Psychology, Professor Tapani Korhonen, who made it possible for me to work with this study. Especially, I wish to thank Professor Heikki Lyytinen who has given me guidance and support through my adventures in the field of science.

Throughout this process my family has been a tremendous source of support and I feel fortunate to be close to my three brothers, five sisters as well as their families. I also wish to thank my wife's family for their friendship. Especially, I am grateful for my fantastic eighteen-year-old son Jarkko, with whom I have enjoyed each and every day of my fatherhood. I also wish to thank his mother, my ex-wife, and her family for their friendship. I will address my deepest gratitude to my loving wife Kaisa for our day-to-day life. At present, our cuties, twins Aini and Otso, surprise our family with their amazing capability to learn new things. I am sure that our four-month-old daughter tries to make me laugh. She first shows an expression of surprise and then a smile, repeating this many times. Thank you all for being there.

# Contents

1. Introduction .....	1
2. Latent growth curve models .....	5
2.1. Linear latent growth model.....	6
2.2. Building latent growth models .....	9
2.2.1. Identification of latent growth model parameters.....	9
2.2.2. Estimation of latent growth model .....	11
2.2.3. Testing and diagnostics of latent growth model.....	13
3. Latent growth mixture modeling.....	14
3.1. Latent growth mixture model .....	14
3.2. Building latent growth mixture model.....	15
3.2.1. Estimation of latent growth mixture model.....	16
3.2.2. Evaluating the number of latent classes of latent growth mixture model.....	20
4. Results of previous simulation studies .....	23
4.1. Latent growth model simulation studies.....	24
4.2. Mixture model simulation studies .....	25
5. Monte Carlo simulation study .....	29
5.1. Monte Carlo method .....	29
5.2. Research questions and model under investigation.....	30
5.3. Implementation of the simulation study .....	36
5.4. Problems in estimation of latent growth mixture model .....	37
5.5. Deciding on the number of latent classes .....	38
5.6. Criteria used to evaluate parameter estimation.....	39
6. Results of the simulation study .....	43
6.1. Problems in estimation of latent growth mixture model .....	44
6.1.1. Pilot simulation study .....	44
6.1.2. Problems in estimation of latent growth mixture model in the main simulation study .....	47
6.2. Results of deciding the number of latent classes.....	51
6.2.1. Results of comparing two-class solution versus one-class solution.....	51
6.2.1.1. Results of AIC, BIC and aBIC .....	51
6.2.1.2. Results of VLMR, LMR and BLRT.....	62
6.2.1.3. Evaluating information criteria to produce wrong two- class solution versus right one-class solution.....	69
6.2.1.4. Summary of the results comparing two-class solution versus one-class solution.....	72

6.2.2.	Results of the wrong three-class solution versus the right two-class solution for model A.8 .....	75
6.2.2.1.	Results of AIC, BIC and aBIC .....	75
6.2.2.2.	Results of VLMR, LMR, BLRT and OLRT .....	76
6.2.2.3.	The effect of a larger number of starting values .....	77
6.2.2.4.	Evaluating the information criteria to produce the wrong three-class solution versus the right two-class solution .....	79
6.3.	Results of evaluation of parameter estimation .....	80
6.3.1.	Results of MSE .....	80
6.3.1.1.	Results of MSE for intercept parameters $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$ ..	81
6.3.1.2.	Results of MSE for slope parameters $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$ .....	90
6.3.1.3.	Results of MSE of variances $\psi_{00}$ , $\psi_{11}$ and covariance $\psi_{01}$ parameter estimation.....	98
6.3.1.4.	Results of MSE of error variances $\theta_1, \theta_2, \theta_3$ and $\theta_4$ estimation .....	107
6.3.1.5.	Summary of the results of MSE .....	111
6.3.2.	Results of proportion of parameter bias in MSE (PB).....	115
6.3.2.1.	Results of PB for $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$ .....	115
6.3.2.2.	Results of PB for $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$ .....	118
6.3.2.3.	Results of PB for $\psi_{00}$ , $\psi_{11}$ and $\psi_{01}$ .....	120
6.3.2.4.	Results of PB for $\theta_1, \theta_2, \theta_3$ and $\theta_4$ .....	126
6.3.2.5.	Summary of the results of PB .....	128
6.3.3.	Results of relative bias of asymptotic standard error (RB) .....	129
6.3.3.1.	Results of RB for $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$ .....	129
6.3.3.2.	Results of RB for $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$ .....	136
6.3.3.3.	Results of RB for $\psi_{00}$ , $\psi_{11}$ and $\psi_{01}$ .....	143
6.3.3.4.	Results of RB for $\theta_1, \theta_2, \theta_3$ and $\theta_4$ .....	151
6.3.3.5.	A summary of the results of RB .....	154
6.3.4.	Results of 95 % coverage for parameters .....	156
6.3.4.1.	Results of 95 % coverage for $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$ .....	156
6.3.4.2.	Results of 95 % coverage for $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$ .....	161
6.3.4.3.	Results of 95 % coverage for $\psi_{00}$ , $\psi_{11}$ and $\psi_{01}$ .....	166
6.3.4.4.	Results of 95 % coverage for $\theta_1, \theta_2, \theta_3$ and $\theta_4$ .....	173
6.3.4.5.	Summary of results of 95 % coverage .....	176
6.3.5.	Results of estimated class proportion .....	178
7.	Discussion .....	180
7.1.	Results of previous studies and their limitations .....	181
7.2.	Conclusions based on the pilot study .....	182
7.3.	Conclusions based on the main simulation study .....	182
7.3.1.	Deciding the number of latent classes .....	182

7.3.2. Failed estimation and number of negative variance estimates .....	183
7.3.3. Results of evaluation of parameter estimation .....	184
7.3.3.1. Results of MSE.....	184
7.3.3.2. Results of proportion of bias in MSE.....	184
7.3.3.3. Results of relative bias of asymptotic standard error ..	184
7.3.3.4. Results of 95 % coverage .....	185
7.4. Results of the simulation study that should be accounted for in empirical research .....	186
7.5. Implications for further studies.....	188
References .....	190
Appendices .....	198

# 1. Introduction

The recent development of latent growth modeling (LGM) provides a better tool to model development in longitudinal data than older statistical methods. The theory of LGM goes back two decades when McArdle and Epstein (1987), and Meredith and Tisak (1990), suggested a theory of latent growth modeling based on covariance matrix and mean vector. This model consisted of estimated growth function with fixed and random parts, which describes the average development in the population, as well as the variation of individual development, respectively (Meredith & Tisak 1990).

This modeling idea has been extended to include the estimation of the impact of covariates on individual growth (McArdle & Aber, 1990) and possibilities to use - besides observed variables with continuous scale - categorical variables (Muthén, 2004). The estimation of model parameters can be managed using structural equation modeling (SEM) programs, such as LISREL (Jöreskog et.al., 1999), EQS (Bentler, 1995), Mx (Neale et al., 1999), Amos (Arbuckle, J.L., 2006) and Mplus (Muthén & Muthén, 1998-2006). Modeling development in the SEM context opens up possibilities to test and estimate the growth, or development, of both groups and individuals at the same time, and also to relate the development to other variables (see for example, Aunola, Leskinen, Onatsu-Arvilommi, & Nurmi, 2002; Chan, Ramey, Ramey & Schmitt, 2000; Curran & Hussong, 2003, Duncan, Duncan, Strycker, Li & Alpert, 1999; Mason, 2001, Tolvanen, 2000).

Another important recent modeling extension within the SEM context is mixture modeling. This modeling is based on the idea that observed data can represent subpopulations, i.e. latent classes, those classes can be identified and their parameters estimated (Muthén & Shedden, 1999; Muthén, 2001). The distribution of observed variables is through mixed distribution, so that each subpopulation has its own model parameter values (see for example, Lubke, Muthén, 2005; Lyytinen, Tolvanen, Torppa, Poikkeus, & Erskine, 2006; Muthén, 2006; Torppa, Tolvanen, Poikkeus, Eklund, Lerkkanen, Leskinen, & Lyytinen, in press).



A combination of these two above mentioned possibilities to analyze data, i.e. latent growth curve modeling combined with the idea of latent classes, is the method under investigation in this study (Muthén, 2001; Muthén, 2004). As noted in the paper of Bauer and Curran (2003b), it is not surprising that this dynamic way of analyzing longitudinal data takes an increasingly central position in the social sciences, e.g. in psychology (Fuzhong, Barrera, Hops, Fisher, 2002; Parrila, Aunola, Leskinen, Nurmi & Kirby, 2005; van Lier, Muthén, van der Sar & Crijnen, 2004).

Meredith and Tisak (1990) and Chou, Bentler and Pentz (1998) have used the term 'latent curve analysis' (LCA), whereas Duncan et al. (1999) and Muthén (2004) have used the term 'latent growth curve model' to refer to modeling growth in the context of SEM. In this study, the term 'latent growth curve model' (LGM or LG model) is used. The term 'latent growth mixture modeling' (LGMM or LGM modeling; Muthén, 2001; Muthén, 2004), in turn, is used to refer to the combination of latent growth curve modeling and mixture modeling.

The conventional latent growth modeling is based on the idea that observed data consists of one population. The LGM then consists of an average growth pattern defined by the mean value of the latent intercept component as well as the mean value of the latent slope component, on one hand, and an individual variation around this pattern, defined by the variance of latent intercept and slope components, on the other (e.g., Bollen & Curran, 2006). However, another possibility is that observed data consists of different subpopulations. Here, the distribution of the observed variables is a mixed distribution, so that each subpopulation has its own model parameter values.

For example, the LGM parameter values for the mean of intercept and slope components differ between subpopulations. The latent growth curve modeling combined with the idea of latent classes, i.e. latent growth mixture modeling (LGMM or LGM modeling; Muthén, 2001, 2004), makes it possible to identify and estimate these subpopulations. In this way, data can be simultaneously analyzed through both variable and person-oriented ways. Taking account of both of these approaches provides a more holistic view in social sciences (Bergman, Magnusson & El-Khoury, 2003). After finding latent classes each observation has probabilities to belong to each latent class and these probabilities provide bases to extend the modeling, either to predict the latent class membership or to relate the latent classes to outcome variables.

Despite twenty years development of the theory behind the LGM and LGMM, these are novel methods in analyzing data in practice. As these methods become

more common and important to model development, the functionality of the models needs more investigation (Bauer & Curran, 2003a, 2003b; Muthén, 2004). Because the theory of the models is based on asymptotic results, researchers can trust the results when sample size is large, e.g. over 1000. However, in many empirical studies, sample size is limited to only 100-500 cases. Simulated data makes it possible to examine functionality of these methods also with small sample sizes.

The goal of this study is to examine the functionality of the LGMM through a limited number of observations. This examination is carried out by producing data consisting of random observations, which are based on prior defined parameter values of the LGMM and by analyzing this simulated data with the LGMM. By repeating a large number of simulations and gathering information from these together, makes it possible to draw conclusions concerning the functionality of the LGMM with different sample sizes. The differences in parameter values between subpopulations obviously have an effect on the functionality of the LGMM with smaller sample sizes. Consequently, the effect of this is examined in this study. Other affecting factors examined are the reliability of observed variables, number of repeated measures, and model construct.

In Chapter 2, the theory of LGM is first concisely explained. Section 2.1 consists of the properties of the used linear LGM. In section 2.2, the LGM is discussed in the following stages used, when proceeding in the LGM: identification of the LGM, estimation of the LGM, testing the model fit, and studying the diagnostics of the parameters of the LGM.

Second, in Chapter 3, the LG model is extended to the mixed distribution with an unknown number of subpopulations. Observed distribution is mixed in the sample, within which the number of classes and class proportions are unknown. This requires the use of an EM algorithm to estimate the model parameters (as will be explained in section 3.2.1), and to evaluate the number of groups (as will be explained in section 3.2.2).

Third, in Chapter 4, previous simulation studies carried out are described. In section 4.1, the simulation studies examining the LGM are introduced and in section 4.2., a few simulation studies related to the LGMM are described. The fact that there are only a few previous simulation studies on the LGMM, makes simulation studies in this study a more demanding case.

In Chapter 5, the Monte Carlo simulation study used in this dissertation is introduced by describing the simulation method in section 5.1, the research

question and model under investigation in section 5.2, implementation of simulation study in section 5.3, and the information of used indicators to evaluate the simulation results in sections 5.4 – 5.6.

In Chapter 6, the results of the simulation study carried out in this study are described, consisting of three main parts. First, in section 6.1, the results concerning problems in the estimation of the LGMM are expressed. The results of the pilot simulation study described in subsection 6.1.1, show the boundaries for the functionality of the LGMM in terms of successful estimation and improper results of estimation. These boundaries advise to specify the LGMM parameters in the further simulation study of this work. In subsection 6.1.2, the results of these further examined LGM models concerning unsuccessful estimation and negative variance estimates are presented. Second, section 6.2 consists of the results related to deciding the number of latent classes. Third, section 6.3 consists of the results concerning the evaluation of the parameter estimation.

Finally, Chapter 7 consists of the conclusions of the present simulation study and their consequences, which should be taken into account, for example, in empirical studies applying the LGMM.

## 2. Latent growth curve models

The latent growth curve model (LGM) is often composed of two latent components. The first one relates to the level (i.e., intercept) and the other one to the slope of growth. These two factors consist of the mean value of intercept and slope and individual random variation of these two latent components. Slope is either fixed to describe linear change or, alternatively, the pattern of slope is estimated (Aunola et. al., 2004; McCallum, Kim, Malarkey & Kiecolt-Glaser, 1997; Duncan & Tidsley, 1995; Wickrama & Lorenz 1997; Duncan, Duncan, Alpert, Hops, Stoolmiller & Muthén, 1997). It is also possible to use a higher order polynomial latent growth model (see for example Reynolds, Finkel, McArdle, Gatz, Berg, & Pedersen, 2005, Windle & Windle, 2001). When comparing the LGM with the confirmatory two-factor model, the difference between these two models is in unknown coefficients of loadings. The confirmatory factor model usually relates each variable to only one factor when variable's other loadings are fixed to zero. In the LGM, most of the variables are related to both growth factors (i.e., intercept and slope) and are usually fixed to some specific values.

Growth curve analysis has a long history and the contemporary basis of latent growth analysis can be found in the development of multilevel modeling (Goldstein, 1987, 1996), hierarchical linear modeling (HLM, Bryk & Raudenbush 1987, 1992), random effects modeling (Rovine & Molenaar 1998, Longford, 1993) and mixed-effects modeling (MacCulloch & Searle, 2001). These statistical methods have some features that make it suitable for some data structures (for example varying measurement points of individuals) and are favorable in some areas of research, for example, in econometrics, biometrics or human behavior. Linear LGM in section 2.1. can estimate using all of the above methods and yield comparative results. For example, the comparison of using HLM- and LGM-methods yielded the same results when estimating the same function of growth (Chou, Bentler & Pentz 1998). Although in their article McArdle and Nesselroade (2002) wrote “the term latent growth models seem appropriate for any technique that describes the underlying growth in terms of latent changes using classical assumptions (e.g., independence of residual errors)”, the term LGM referred to in this study means that the latent growth model is defined within the structural equation context.

An LGM consisting of only a single variable for each repeated measure can extend to latent components with many indicators in each measurement time (see Anstey, Hofer, Luszcz, 2003, Hancock, Kuo, Lawrence, 2001, Tolvanen, 2000). In this study, an LGM with one indicator for each repeated measure is used.

## 2.1. Linear latent growth model

In the LG modeling, an average development over time and individual variation around this average are of interest. The pattern of slope component is the same for each individual, but the strength of this pattern may vary individually. The following consists of the basic theory of the LGM. In this work, the slope component of the LGM is linear and the comparison time point is the first one.

The latent growth model with linear slope component is defined with two parts:

I Measurement part

$$y_{it} = \eta_{0i} + (t-1)\eta_{1i} + \varepsilon_{it}, \quad i = 1,2,\dots,n \text{ and } t = 1,2,\dots,T, \quad (2.1)$$

where

$y_{it}$  is the observation of individual  $i$  at the time point  $t$

$\eta_{0i}$  is the intercept component of individual  $i$ ,

$\eta_{1i}$  is the linear slope component for individual  $i$ ,

$\varepsilon_{it}$  is the measurement error for individual  $i$  at the time point  $t$ ,

$n$  is the number of observations,

$T$  is the number of measurements.

II Latent part

$$\begin{cases} \eta_{0i} = \alpha_0 + \zeta_{0i}, \\ \eta_{1i} = \alpha_1 + \zeta_{1i}, i = 1, 2, \dots, n, \end{cases} \quad (2.2)$$

where

$\alpha_0$  and  $\alpha_1$  are expected values (fixed part of the model) of latent growth factors  $\eta_0$  and  $\eta_1$ , respectively,

$\zeta_0$  and  $\zeta_1$  are random variables that configure individual growth.

Denote the covariance matrix of  $\boldsymbol{\varepsilon}$  and  $\boldsymbol{\eta} = (\eta_0, \eta_1)^T$  as follows

$$\text{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta} = \text{diag}(\theta_1, \theta_2, \dots, \theta_T)$$

and

$$\text{cov}(\boldsymbol{\eta}) = \boldsymbol{\Psi} = \begin{bmatrix} \psi_{00} & \psi_{01} \\ \psi_{01} & \psi_{11} \end{bmatrix}, \quad (2.3)$$

respectively.

The model in this study consists of the first two polynomial functions, but it is possible to define the model by using higher order polynomials as well (see Bollen & Curran, 2006; Tolvanen, 2000). In the above LGM, repeated measures are defined to be of equal interval. This interval can vary along successive measurements and along individuals. In the LG model, coding some time point to zero fixes the comparison of growth to that time point (Biesanz, Deeb-Sossa, Papadakis, Bollen & Curran, 2004).

The latent linear growth model (2.1) can be presented by using a general SEM framework;

$$\begin{cases} \mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{\zeta} \end{cases}, \quad (2.4)$$

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & T-1 \end{bmatrix},$$

where

$\mathbf{y}$  is a  $T \times 1$  vector of the observed variables,

$\boldsymbol{\eta}$  is a  $2 \times 1$  vector of the latent components,

$\boldsymbol{\varepsilon}$  is a  $T \times 1$  vector of measurement errors,

$\boldsymbol{\alpha}$  is a  $2 \times 1$  vector of expected values of  $\boldsymbol{\eta}$ .

The expected values of measurement errors  $\boldsymbol{\varepsilon}$  are zeros and covariance matrix of  $\boldsymbol{\varepsilon}$  is denoted by  $\boldsymbol{\Theta}$ .

Assuming that  $E\boldsymbol{\varepsilon} = \mathbf{0}$  and  $E\boldsymbol{\zeta} = \mathbf{0}$  the expectation vector and the covariance matrix of the latent components  $\boldsymbol{\eta}$  are

$$E\boldsymbol{\eta} = \boldsymbol{\mu}_{\boldsymbol{\eta}} = \boldsymbol{\alpha},$$

$$\text{cov}(\boldsymbol{\eta}) = \boldsymbol{\Psi},$$

and the expectation vector and the covariance matrix of observed variables  $\mathbf{y}$  are

$$E\mathbf{y} = \boldsymbol{\mu}_y = \Lambda E(\boldsymbol{\eta}) = \Lambda \boldsymbol{\alpha}, \quad (2.5)$$

$$\text{cov}(\mathbf{y}) = \boldsymbol{\Sigma} = \Lambda \boldsymbol{\Psi} \Lambda^T + \boldsymbol{\Theta}, \quad (2.6)$$

respectively.

For example, when  $T = 3$ ,

$$E\mathbf{y} = \begin{bmatrix} \alpha_0 \\ \alpha_0 + \alpha_1 \\ \alpha_0 + 2\alpha_1 \end{bmatrix},$$

and

$$\text{cov}(\mathbf{y}) = \begin{bmatrix} \psi_{00} + \theta_1 & & \\ \psi_{00} + \psi_{01} & \psi_{00} + 2\psi_{01} + \psi_{11} + \theta_2 & \\ \psi_{00} + 2\psi_{01} & \psi_{00} + 3\psi_{01} + 2\psi_{11} & \psi_{00} + 4\psi_{01} + 4\psi_{11} + \theta_3 \end{bmatrix}.$$

## 2.2. Building latent growth models

Building an LGM is similar to the traditional structural equation modeling. The distinguishable stages in modeling are 1) specification of the model, 2) checking the identifiability of the model, 3) estimation of the model, 4) testing and evaluating overall fit of the model, and 5) doing diagnostics of parameters of the model. If the model fits the data poorly, parameters are non-significant, or the model needs some modification, it is possible to re-specify the model or to test alternative models. In this case, the process is repeated through the stages 1-5.

When the model is specified as a linear LGM in equation 2.1 and 2.2, individual growth is modeled with a very parsimonious model consisting of two parameters to describe the overall growth, two parameters to describe individual variation in growth and one parameter to describe covariance between the intercept and slope. One advantage in the LGM is that the measurement error can be isolated by estimating error variances.

### 2.2.1. Identification of latent growth model parameters

In order to get an identifiable model, the model must be identifiable both in the covariance and expectation structure. Because equations of expectation and



covariance parameters do not depend on each other (see equations 2.5 – 2.6), the identifiability of the model can be considered separately for both model parts (Bollen & Curran, 2006).

When modeling expectations of latent factors, there are only two parameters,  $\alpha_0$  and  $\alpha_1$ , to be estimated. If there is, for example, three time points, this leads, according to equation 2.5 to,

$$\begin{aligned}\mu_1 &= \alpha_0 \\ \mu_2 &= \alpha_0 + \alpha_1 \\ \mu_3 &= \alpha_0 + 2\alpha_1.\end{aligned}$$

From these equations, the identification of the mean of the intercept component and the mean of the slope component can be established, for example,

$$\begin{aligned}\alpha_0 &= \mu_1 \\ \alpha_1 &= \mu_2 - \mu_1.\end{aligned}$$

The modeling covariance structure with the linear LGM in equation 2.6 has two parameters for the variances of latent components, one parameter for the covariance between components, and one parameter for the variance of measurement error at each time point. For example, a linear LGM with three repeated measurements have six parameters to be estimated, and in a covariance matrix, there are six equations to resolve these parameters,

$$\begin{aligned}\text{var}(y_1) &= \psi_{00} + \theta_1 \\ \text{cov}(y_1, y_2) &= \psi_{00} + \psi_{01} \\ \text{cov}(y_1, y_3) &= \psi_{00} + 2\psi_{01} \\ \text{var}(y_2) &= \psi_{00} + 2\psi_{01} + \psi_{11} + \theta_2 \\ \text{cov}(y_2, y_3) &= \psi_{00} + 3\psi_{01} + 2\psi_{11} \\ \text{var}(y_3) &= \psi_{00} + 4\psi_{01} + 4\psi_{11} + \theta_3\end{aligned}$$

From these equations, the identification of the variances and covariance of latent components can be established by,

$$\begin{aligned}\psi_{00} &= 2 \times \text{cov}(y_1, y_2) - \text{cov}(y_1, y_3) \\ \psi_{01} &= \text{cov}(y_1, y_3) - \text{cov}(y_1, y_2) \\ \psi_{11} &= [\text{cov}(y_2, y_3) + \text{cov}(y_1, y_2) - 2 \times \text{cov}(y_1, y_3)] / 2.\end{aligned}$$

The identification of error variances can then be established by,

$$\begin{aligned}\theta_1 &= \text{var}(y_1) - \psi_{00} \\ \theta_2 &= \text{var}(y_2) - \psi_{00} + 2\psi_{01} + \psi_{11} \\ \theta_3 &= \text{var}(y_3) - \psi_{00} + 4\psi_{01} + 4\psi_{11}.\end{aligned}$$

In order to have an identifiable linear LG model, three measurements are necessary and sufficient. However, to be able to test that a linear LG model appropriately fits the covariance structure, at least four measurements are required.

### 2.2.2. Estimation of latent growth model

Suppose that in the LGM defined in equations 2.4, latent components and errors are normally distributed

$$\begin{aligned}\zeta &\sim N_2(\mathbf{0}, \mathbf{\Psi}), \\ \eta &\sim N_2(\boldsymbol{\alpha}, \mathbf{\Psi}), \\ \boldsymbol{\varepsilon} &\sim N_T(\mathbf{0}, \mathbf{\Theta}).\end{aligned}\tag{2.7}$$

Then the observations of  $T \times 1$  vector  $\mathbf{y}$  according equations 2.5-2.7 are multi-normally distributed

$$\mathbf{y}_i \sim N_T(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad i = 1, 2, \dots, n,$$

where

$$\boldsymbol{\mu} = \boldsymbol{\Lambda}\boldsymbol{\alpha}$$

and

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}^T + \boldsymbol{\Theta}.$$

Denote  $\boldsymbol{\tau} = \text{vec}(\boldsymbol{\alpha}, \boldsymbol{\Psi}, \boldsymbol{\Theta})$  is  $P \times 1$  vector, which consists of all free parameters in  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\Psi}$  and  $\boldsymbol{\Theta}$ . The parameters  $\boldsymbol{\tau}$  can be estimated using the maximum likelihood estimation (ML), whose log-likelihood can be expressed as

$$\log L = -c - (n/2) \log |\boldsymbol{\Sigma}(\boldsymbol{\tau})| - (1/2)A, \quad (2.8)$$

where

$$c = nT / [2 \log(2\pi)]$$

and

$$A = \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}) (\mathbf{y}_i - \boldsymbol{\mu}) = n \times \text{tr} \left[ \boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}) (\mathbf{S} + (\bar{\mathbf{y}} - \boldsymbol{\mu})(\bar{\mathbf{y}} - \boldsymbol{\mu})^T) \right],$$

where

$\bar{\mathbf{y}}$  is the sample mean vector and

$\mathbf{S}$  is the sample covariance matrix.

The theoretical model is compared with an unrestricted model that has a log-likelihood function,

$$\log L_u = -c - (n/2) \log |\mathbf{S}| - (n/2)T. \quad (2.9)$$

The minimization of the likelihood ratio of  $\log L$  and  $\log L_u$  models, the fit function

$$\begin{aligned} F_{ML}(\boldsymbol{\tau}) &= -\log L / n + \log L_u / n \\ &= 1/2 \left\{ \log |\boldsymbol{\Sigma}(\boldsymbol{\tau})| + \text{trace}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}) (\mathbf{S} + (\bar{\mathbf{y}} - \boldsymbol{\mu})(\bar{\mathbf{y}} - \boldsymbol{\mu})^T)) - \log |\mathbf{S}| - T \right\} \end{aligned} \quad (2.10)$$

produces ML estimates for parameter vector  $\boldsymbol{\tau}$ .

Asymptotical standard errors of parameter estimators are on the diagonal of inverse of approximated Fisher information matrix,

$$I_{ML} = - \sum_{i=1}^n \frac{\partial^2 \log L}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T}. \quad (2.11)$$

### 2.2.3. Testing and diagnostics of the latent growth model

Investigating the sufficiency of a specified linear LGM is based on the hypotheses:

H<sub>0</sub>: specified LG model fit to the data sufficiently

H<sub>1</sub>: unrestricted model.

If the null hypothesis is true, the estimated fit function of model,  $F_{ML}(\boldsymbol{\tau})$ , multiplied with two times the number of observations is distributed as chi square

$$2nF_{ML}(\hat{\boldsymbol{\tau}}) \sim \chi^2(df), \quad (2.13)$$

where

$$df = (T^2 + T)/2 + T - P .$$

Another way to evaluate the overall fit of the model is to use various fit indices. If the data consists of a large number of observations, the  $\chi^2$ -test is usually, in practice, statistically significant even when the fit of the model to the data is good enough from a practical point of view. Alternative fit indices that can be used alongside the  $\chi^2$ -test are, for example, the root mean square error of approximation (RMSEA), Tucker Lewis index (TLI), comparative fit index (CFI), and standardized root mean square residual (SRMR). The model fits the data well if RMSEA is lower than .06, TLI and CFI are higher than .95, and SRMR is lower than .08 (Hu & Bentler, 1999).

If the fit of the model is poor, alternative models can be tested or the model can be modified on the basis of the diagnostics of parameters. Checking t-values of parameters identifies those parameters that are not statistically significant. These non-significant parameters are then fixed to zero. The other possibility to modify the model is to add some parameters into the model. The parameters that should be added can be found with the help of modification indices. In both of the above mentioned cases, the modified model should be estimated over again. After modifying the model, the process is repeated through stages 1-5.

### 3. Latent growth mixture modeling

In Chapter 2, the data was supposed to consist of observations from one population. The theory and model building of LGM can now be extended to two or more subpopulations, where the number of subpopulations and members of these subpopulations are unknown beforehand (see McLahlan & Peel, 2000; Yung, 1997). The collected data is one sample, consisting of many subpopulations with unknown proportions, and therefore, the distribution of observations is a mixture of distributions of many subpopulations. The parameters for the decided number of latent classes can be estimated from mixture data using EM algorithm (see Muthén & Shedden, 1999; McLahlan & Krishnan, 1997). This theory is also appropriate for estimating a latent growth mixture model.

#### 3.1. Latent growth mixture model

Suppose the linear LG model for each subpopulation  $k$ . Then the model is the following form

$$\begin{cases} \mathbf{y}^{(k)} = \mathbf{\Lambda} \boldsymbol{\eta}^{(k)} + \boldsymbol{\varepsilon}^{(k)} \\ \boldsymbol{\eta}^{(k)} = \boldsymbol{\alpha}^{(k)} + \boldsymbol{\zeta}^{(k)}, \end{cases} \quad (3.1)$$

where

$$\boldsymbol{\varepsilon}^{(k)} \sim N(0, \boldsymbol{\Theta}^{(k)})$$

and

$$\boldsymbol{\zeta}^{(k)} \sim N(0, \boldsymbol{\Psi}^{(k)}), k=1, \dots, K.$$

Possible differences between classes are the differences in expectation vectors  $\boldsymbol{\alpha}^{(k)}$  of latent components, covariance matrices  $\boldsymbol{\Psi}^{(k)}$  of latent components or covariance matrices of measurement errors  $\boldsymbol{\Theta}^{(k)}$ . The most interesting differences between classes are the differences in expectations of latent components, which differences mean different developmental trajectories between classes.

When estimating the LGMM parameters, there are additional parameters compared to the LGM, namely proportions  $p_k$  of latent classes  $k=1,2,\dots,K$ . The number of these free parameters is  $K-1$ , because they have the restriction

$$\sum_{k=1}^K p^{(k)} = 1.$$

### 3.2. Building latent growth mixture model

Building a LGMM is a closely similar process to the building of the LGM, with the difference of deciding the number of latent classes, instead of testing and evaluating the overall fit of the model. The distinguishable stages in the LGM modeling are 1) specification of the model, 2) checking identification of the model, 3) estimation of the model for 1,2,...,K class solutions, 4) testing the right number of latent classes and 5) doing diagnostics of parameters.

#### *Specification*

Specification of LGMM can be started by fixing covariance matrix equal between classes  $\boldsymbol{\Psi}^{(k)} = \boldsymbol{\Psi}$  and fixing the error covariance matrix of measurement errors equal between classes  $\boldsymbol{\Theta}^{(k)} = \boldsymbol{\Theta}$ ,  $k = 1,2,\dots,K$ . These ensure that the model is simple enough to be empirically identifiable and estimable. By using modification indices as a guide, the excess constraints can be found and freed. Sometimes the covariance matrices of latent factors are also fixed to zero  $\boldsymbol{\Psi}^{(k)} = 0$ ,  $k = 1,2,\dots,K$  in which case the differences between individuals are in the mean trajectories of classes, and the differences between individuals are supposed to come from random error,  $\varepsilon$  (Nagin, 1999; Jones et al., 2001). This stringent model is shown to have a statistically significantly poor fit against the above model (3.1) in real data (Bauer & Curran, in press).

## *Identification*

After specifying the LGM, the first step is to ensure that the model is also identifiable in the case of several classes. This identifiability can be checked by considering the LGMM parameters, as in the case of the previously presented one-sample LGM. The LGMM also includes parameters for class sizes. The parameters of class sizes are not always identifiable, especially if the model contains more classes than the sample has in reality. In this context, the model is not necessarily empirically identifiable. After estimating models with a different number of latent classes, these models are compared and the decision about the right number of latent classes is made. To ensure the identifiability of LGMM empirically, one has to start from a model that is known to be identified and check if the observed-data log likelihood changes when adding a parameter (Muthén & Shedden, 1999; Muthén & Muthén, 1998-2006).

The next two important steps when building an LGMM, are discussed in their own sections. The section 3.2.1 contains more detailed information concerning the estimation of LGMM, and section 3.2.2 discusses the evaluation of number of latent classes further.

## *Diagnostics of parameters*

Parameters that are not statistically significant are identified by checking the *t*-values of the parameters. These non-significant parameters are then fixed to zero. The other possibility to modify the model is to free some constraints between classes, or to add some parameters into the model with the help of modification indices. In both of the above mentioned cases, the modified model should be estimated over again and, therefore, the process is repeated through stages 1-5.

### **3.2.1 Estimation of latent growth mixture model**

In the LGMM, the parameter estimation using the ML estimation method is implemented through an EM algorithm (Muthén & Shedden, 1999; McLahlan & Krishnan, 1997; Yung, 1997). The estimation of LGMM consists of two parts: the estimation of parameters related to the LGM and the estimation of class proportion. Thus, the log-likelihood function of observed data for the LGMM is

$$\log L = \log \prod_{i=1}^n L_i = \sum_{i=1}^n \log L_i = \sum_{i=1}^n \log f(\mathbf{y}_i), \quad (3.2)$$

where density function  $f$  is mixed from  $K$  density functions

$$f(\mathbf{y}) = \sum_{k=1}^K p^{(k)} f^{(k)}(\mathbf{y}), \quad (3.3)$$

where

$p^{(k)}$  is the proportion of subpopulation  $k$  in the population.

The density function for class  $k$  is

$$f^{(k)}(\mathbf{y}) \sim N_T(\boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}), \quad (3.4)$$

where

$$\begin{aligned} \boldsymbol{\mu}^{(k)} &= \boldsymbol{\Lambda}^{(k)} \boldsymbol{\alpha}^{(k)}, \\ \boldsymbol{\Sigma}^{(k)} &= \boldsymbol{\Lambda}^{(k)} \boldsymbol{\Psi}^{(k)} \boldsymbol{\Lambda}^{(k)T} + \boldsymbol{\Theta}^{(k)}. \end{aligned}$$

Denote the class information with the vector  $\mathbf{c}_i = (c_{i1}, c_{i2}, \dots, c_{iK})^T$ , where  $c_{ik} = 1$ , if observation  $i$  belongs to class  $k$  and  $c_{ik} = 0$  otherwise.

The conditional density function is

$$f(\mathbf{y}_i | \mathbf{c}_i) = \sum_{k=1}^K P(c_{ik} = 1) f_k(\mathbf{y}_i | c_{ik} = 1), \quad (3.5)$$

where

$$\begin{aligned} f^{(k)}(\mathbf{y}_i | c_{ik}) &= f^{(k)}(\mathbf{y}), \\ P(c_{ik} = 1) &= p^{(k)}, \end{aligned}$$

so that

$$\sum_{k=1}^K p^{(k)} = 1.$$



When the class information  $c_{ik}$  is known, the complete-data log likelihood is

$$\begin{aligned}
& \log \prod_{i=1}^n f(\mathbf{y}_i | \mathbf{c}_i) \\
&= \log \prod_{i=1}^n \prod_{k=1}^K [P(c_{ik} = 1) f(\mathbf{y}_i | c_{ik})]^{c_{ik}} \\
&= \sum_{i=1}^n \left[ \log \prod_{k=1}^K P(c_{ik} = 1)^{c_{ik}} f(\mathbf{y}_i | c_{ik})^{c_{ik}} \right] \\
&= \sum_{i=1}^n \left[ \log \prod_{k=1}^K P(c_{ik} = 1)^{c_{ik}} + \log \prod_{k=1}^K f(\mathbf{y}_i | c_{ik})^{c_{ik}} \right] \\
&= \sum_{i=1}^n \left[ \sum_{k=1}^K c_{ik} \log P(c_{ik} = 1) + \sum_{k=1}^K c_{ik} \log f(\mathbf{y}_i | c_{ik}) \right].
\end{aligned} \tag{3.6}$$

The above equation consists of two independent parts to maximize the complete-data log-likelihood, namely the sum of the weighted K class probabilities  $P(c_{ik} = 1)$  and the sum of the weighted K density function  $f(\mathbf{y}_i | c_{ik})$  (Muthén & Shedden, 1999).

When maximizing the log-likelihood with EM algorithm, the latent class information  $c_{ik}$  is considered missing. The EM algorithm includes E-step (expectation step) and M-step (maximization step). With E-step, the expected values of observations belonging to each latent class are calculated respecting the starting values at the first step, and the value from M-step in further iterations. These posterior probabilities (from a Bayesian point of view) for observation  $i$  belonging to class  $k$  is calculated with E step using formula

$$p_{ik} = \frac{p^{(k)} f^{(k)}(\mathbf{y}_i)}{f(\mathbf{y}_i)}, i = 1, 2, \dots, n, k = 1, 2, \dots, K. \tag{3.7}$$

These posterior probabilities are then used in M-step when maximizing expected values in equation 3.6. This leads to maximize

$$E \left\{ \sum_{i=1}^n \sum_{k=1}^K c_{ik} \log P(c_{ik} = 1) \right\} = \sum_{i=1}^n \sum_{k=1}^K p_{ik} \log P(c_{ik} = 1), \tag{3.8}$$

resulting in the values of  $p^{(k)}$  parameters in equations 3.3. or 3.6 and to maximize

$$E\left\{\sum_{i=1}^n [c_{ik} \log f(y_i | c_i)] | y_i\right\} = \sum_{i=1}^n \sum_{k=1}^K p_{ik} \log f^{(k)}(y_i), \quad (3.9)$$

resulting in  $f^{(k)}(y)$  in equations 3.3 or 3.6 with the values of  $\Lambda^{(k)}, \alpha^{(k)}, \Psi^{(k)}$  and  $\Theta^{(k)}$ , where  $k = 1, \dots, K$ .

After M-step, the algorithm returns to E-step to calculate new posterior probabilities and then again to M-step. This iteration continues until the convergence criterion related to the complete-data log-likelihood is met.

It is a known feature that the LGMM estimation may often stop to the local maximum of log-likelihood producing biased parameter estimates (Hipp & Bauer, 2006). The solution to get the highest value of likelihood in data is to use many different starting values for parameters (Muthén, 2004).

This estimation method, called MLR (Muthén 1998-2006), produces ML estimates and standard errors which are robust for nonnormality (McLahlan & Peel, 2000),

$$I_{MLR} = I_{ML}^{-1} I_{MLF} I_{ML}^{-1}, \quad (3.10)$$

where

$$I_{MLF} = \sum_{i=1}^n \frac{\partial \log L_i}{\partial \boldsymbol{\tau}} \times \frac{\partial \log L_i}{\partial \boldsymbol{\tau}^T},$$

$$I_{ML} = -\sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T},$$

and  $L_i$  is a value of an observed data log-likelihood for observation  $i$  defined in equation 3.2.

The estimates of standard errors can be calculated using  $I_{ML}, I_{MLF}$  or  $I_{MLR}$ . In this study, the MLR method, which is more robust for smaller sample sizes (Muthén, & Shedden, 1999), is used.

Before the examination of the estimation results, it is important to evaluate the number of latent classes. This can be based on well-known information criteria or statistical tests. Because this is a very important stage in the LGMM, the next chapter, 3.2.2, concentrates on this.

### **3.2.2 Evaluating the number of latent classes of latent growth mixture model**

Testing and evaluating the overall fit is not possible, in the context of the mixture model, as it is in the framework of conventional structural equation models. Also, the use of the likelihood ratio test to evaluate the right number of latent classes is not possible in mixture analyses. This because likelihood ratio test value is not distributed as chi-square with degrees of freedom equal to the difference between the number of parameters under the null and alternative hypotheses. For this purpose the following criteria are proposed: Akaike's information criteria (AIC), Bayes information criteria (BIC), adjusted BIC (aBIC), Vuong, Lo, Mendell & Rubin (VLMR) test and adjusted VLMR (called Lo, Mendell & Rubin, LMR) test. The most recent possibility included in the Mplus program is to use the parametric bootstrapped likelihood ratio test BLRT.

AIC (Akaike's information criteria) (Akaike, 1987) is defined as a function of log-likelihood and number of estimated parameters  $P$

$$AIC = -2 \log L + 2P . \quad (3.11)$$

When comparing two competitive models, the better one is the model that has the lower AIC.

BIC (Bayes information criteria) (Schwartz, 1978) is defined as follows

$$BIC = -2 \log L + P \log(n) . \quad (3.12)$$

Sclove (1987) proposed that the sample size  $n$  replaces with  $(n+2)/24$  resulting aBIC (adjusted BIC):

$$aBIC = -2 \log L + P \log\left(\frac{n+2}{24}\right). \quad (3.13)$$

As in AIC, the lower BIC and aBIC means a better fitting model.

In the case of one population without any subpopulation, minimizing the likelihood ratio in equation 2.10 produces maximum likelihood estimates. This likelihood ratio distributes asymptotically as a  $\chi^2$  distribution, as presented in equation 2.13, and it is called later, an ordinary likelihood ratio test (OLRT). In the LGM modeling, the use of a likelihood ratio requires that the distribution and its degrees of freedom are defined differently than in OLRT. This is because the parameter values are on the boundary of parameter space in the null hypothesis with one class less than in an alternative hypothesis (McLahlan & Peel, 2000). Therefore, Vuong (1989) proposes (as an extension to Whites (1982) theorem, which is based on Kullback & Leibler (1951) information criteria), the VLMR (Vuong, Lo, Mendell & Rubin test) test, in which the number of latent classes are based on hypotheses

$$\begin{aligned} H_0: & \text{ number of latent classes is } k-1 \\ H_1: & \text{ number of latent classes is } k. \end{aligned}$$

The likelihood ratio of models, respecting the above hypotheses, is compared to its theoretical distribution: that is, under the most general regularity conditions, weighted  $\chi^2$  distribution when models are nested or normal distribution when the models are nonnested (Vuong, 1989).

Lo, Mendell and Rubin (2001) proposed that the above described VLMR (Vuong-Lo-Mendell-Rubin) likelihood ratio test should adjust with the numbers of  $P_k$  and  $P_{k-1}$  of freely estimated parameters in  $k$  and  $k-1$  classes, respectively, and sample size. The adjusted test, called the LMR test in this study, is then

$$LMR = \frac{VLMR}{1 + \{(P_k - P_{k-1}) \log n\}^{-1}}.$$

One method of testing the number of latent classes is to use the parametric bootstrapped likelihood ratio test (BLRT). In this method,  $k$  and  $k+1$  class models are estimated in empirical data. Then, the simulated data based on the parameter estimates of  $k$  class model is randomly generated  $R_B$  times. In each of the generated data, maximum likelihood estimates are calculated for  $k$  and  $k+1$  class models resulting in the likelihood ratio of the models. These  $R_B$  replications of data are used to build a distribution of likelihood ratio for which, the likelihood resulting in empirical data, is compared to evaluate the Type I error rate for the null hypothesis of the  $k$  class model (McLahlan & Peel, 2000; Muthén & Muthén, 1998-2006). The rules deciding the number of replications ( $R_B$ ) are designed by ensuring a clear decision at the .05 nominal level. However, the maximum  $R_B$  is as a default at 100 (Muthén & Muthén, 1998-2006).

## 4. Results of previous simulation studies

To better understand the limitations as well as the possibilities of the new analyzing tool, the LGMM, it is necessary to examine some assumptions behind this method. The assumptions presented by Boomsma and Hoogland (2001) for an ordinary factor analysis in the case of continuous observed variables, are relevant for the LGM as well, and partly also to LGMM. These assumptions are:

- a) independently distributed observations
- b) multinormally distributed observed variables
- c) nearly right hypothetical model
- d) estimation based on covariance matrix
- e) large sample size .

If some of the above mentioned assumptions are not met, a suspicion of unbiased parameter estimates and unbiased standard errors, as well as the indices of model fit, is justified. A violation in the assumptions above raises questions about the consequences of a violation to the results and, moreover, to the conclusions.

Because an interpretation of the components of LGM is directly related to the measurement scale, it is necessary to use a covariance matrix instead of correlation matrix in the model estimation that corresponds with the primary scale (see example, Tolvanen, 2000). This correspondence with the scale is also achieved by using raw data. Raw data is also required, if data has missing observations or if analyzing the LGMM. The basic question is; how many observations are needed for a reliable estimation of the model? The next sections describe some previous simulation studies of the LG model (section 4.1), as well as simulation studies of factor mixture and LGM models (section 4.2 and 4.3). They partly describe an answer to this question and give some guidance in designing LGMM simulation studies.

## 4.1. Latent growth model simulation studies

There are many previous simulation studies for the confirmatory factor analysis, in which, for instance, an estimation with small sample size (Bentler & Yuan, 1999) related to a misspecified model (Fan, Thompson, & Wang, 1999), comparing estimators (Olsson, Troye, & Howell, 1999), the robustness of an estimator against non-normality (Beauducel & Herzberg, 2006), are examined. Impacts on the parameter estimate precision or various fit indices are examined. The other effects in confirmatory factor analysis are for example low reliability of observed variables and model misspecification. Two meta-analyses (Hoogland & Boomsma, 1998; Powell & Schafer, 2001) conclude results from about 50 simulation studies, where the focus has been in the robustness of estimators and the likelihood ratio chi-square test. Because of high sensitivity of the chi-square test, to reject the hypothesized model when the sample size is large, the numerous fit indices are developed to assess the overall fit of the model. Two important simulation studies comparing these indices are by Hu and Bentler (1995, 1999).

The results of the simulation studies for the confirmatory factor analysis can be applied to the LGM, which is a special case of confirmatory factor analysis. However, the estimation is supposed to be more efficient in the LGM than in the confirmatory factor analysis. This efficiency, as the requirement of observation per estimated parameters, could be very small (Jackson, 2001). Unfortunately, there are only a few simulation studies in which the LGM is investigated.

In a simulation study with linear and quadratic growth curves (Tolvanen, 2000), simulated data with varying sample sizes  $n = 50, 100$  or  $500$  consisted of four measurements and 500 replications. Reliability of measurements was set low (.50) or high (.90), alternatively. All growth factors (i.e., intercept, linear slope and quadratic curve component) had the same variance, and the intercept and linear slope were set to correlate .33. The models fitted to the data were either right, or some parameter was wrongly fixed to zero or wrongly freely estimated.

The results showed that when fitting the right model, the overall  $\chi^2$ -test worked well, already with the smallest sample size,  $n = 50$ . The parameter estimation was unbiased with a low sample size but standard errors of the parameter estimators were large, producing non-significant results. In the case of low reliability, standard errors were over two times larger for the intercept and over three times larger for the linear curve than in the case of higher reliability. When the sample size increased from 50 to 100, standard errors of parameters decreased approximately to 70 % from the initial value. When the sample size increased from

50 or 100 to 500, standard errors of parameters decreased approximately to 31 % or 44 % from the initial value, respectively.

Fixing the variance of the quadratic term wrongly to zero, the overall  $\chi^2$ -test rejected the model in all cases, in which the model's sample size was 500 and reliability of observed variables were .90. This wrongly specified model produced biased parameter estimates and standard errors increased substantially. For example, the standard error of variance estimation for the intercept increased by about 5 times. Indeed, if the model is wrongly specified, setting the covariance between the intercept and slope to zero, the model was accepted with the  $\chi^2$ -test approximately in 65 % of the samples in nominal  $p=.05$  level. If this model is compared with the right specified model, the  $\chi^2$ -difference test rejects wrongly specified models in 53 % of the samples.

When the model consists of freely estimated parameters, whose true values are zero, most of the parameter estimates are unbiased but standard errors increase. This addition of parameters has no effect to the  $\chi^2$ -test, as expected.

In empirical studies with small sample size, some parameters are non-significant, and therefore, fixed to zero. This can lead to an underparameterized model, whose parameter estimates are biased and their standard errors increase even two times larger than when compared with the right model.

## **4.2. Mixture model simulation studies**

There are only a few simulation studies carried out in the context of the factor mixture model or latent growth mixture model. First, Lubke and Muthén (2007) reported their results of a mixture model simulation study consisting of two latent classes. The examined models in their study were latent profile analysis with eight observed variables and three different factor models, where the number of factors varied from one to three with 4-8 observed variables. The results showed the extent to which observations were classified into the right classes and in how many cases the estimated 95 % confidence interval consisted of the true parameter value. Lubke and Muthén (2007) evaluated average class probabilities and, more extensively, how entropy measure works as an indicator of the correct class assignment. In the study, the models were examined with three distances between the latent class means, that is, MD is 1, 1.5 or 2 (definition of MD, Mahalanobis



distance, is presented in Chapter 6). Class probabilities were .5 for each of two latent classes, and 120 repeated samples with sample size  $n = 300$  were generated.

Lubke and Muthén (2007) concluded that the parameter coverage of the factor mixture model are good, even for small class separation, whereas the correct class assignment is satisfactory only when the classes are well separated. The complexity of the within-class model with respect of the factor structure, or the number of observed variables within class, does not seem to greatly influence model performance. Further, Lubke and Muthén (2007) also pointed out that the covariates, that predict class membership, are important to include in the model when examining the number of latent classes. This is obvious, because it is comparable to a situation where groups are known and, therefore, the power of test is strong. One important proposition is to use a two-step analysis where loadings are estimated in the first step and fixed to the estimates in the second step. The conclusion is that fixed factor loadings can be a considerable improvement to the model performance, which is the case when fitting the LGMM.

Another simulation study carried out by Nylund, Asparouhov and Muthen (in press) compared the statistical indicators to resolve the number of latent classes for the LGMM and for the latent class analysis (LCA) with continuous and categorical variables. The examined statistical indicators consisted of four information criteria, i.e., AIC, BIC, aBIC and CAIC ( $CAIC = BIC + P$  where  $P$  is the number of freely estimated parameters), and three statistical tests, i.e., LMR, BLRT and ordinary likelihood ratio test (OLRT). The theory for the OLRT requires that compared models are nested, which in the case of mixture models, is not met.

For the linear LGMM, there were four indicators with three latent classes, whose true sizes were 18, 29 and 53 percents. The results of the simulation study for this model were based on 100 replications. The low replication number was due to the long computing time when calculating the BLRT test value. For the LCA, three different models for continuous variables were examined. First, a model with 8 items and 4 latent classes, whose proportions were equal, was examined. Second, a model with 15 items with equal class sizes was examined. For these two models, the number of replications was 100. Third, a model with 10 items and 4 latent classes was examined. The class proportions for four classes were .05, .10, .15 and .75. Because of the smallest class sizes, replication for this model was 500, in order to get more reliable results. The examined sample sizes were  $n = 200, 500$  or 1000. Despite of the small number of replications and few examined statistical models, the results of the simulation study carried out by Nylund et al. revealed important information.

First, the results revealed that the 95 % coverage for all of the models, with all of the examined sample sizes, were found to be very good the 95 % coverage being between .92 and .98, with one exception. The exception was a 10-item categorical LCA model, in which the coverage was .54, .79 and .91 when sample size was 200, 500 and 1000, respectively. This result is obvious, because the expected numbers of cases for the smallest class sizes are 10, 25 and 50 cases with those sample sizes (5 percent of cases). These results point out that the estimation of parameters and their standard errors are successfully measured by the coverages in these models.

Second, examination of Type I error at .05 nominal level revealed that the BLRT test behaved very well producing .02 - .06 error rates for the examined models. The LMR test behaved quite well for the LCA model with continuous variables producing .02-.06 error rates, except the 8-item model with a sample size of 200, in which case the error rate was .11. For the LGM model, the LMR test produced .06, .12 and .25 error rates, when sample sizes were 200, 500 and 1000, respectively. These results for the LMR test mean that the behavior of the test is dependent on the examined model. As can be assumed, the error rates for OLRT seem to be large (.13 - .99) which warns against using this test when aiming to find the right number of latent classes.

Third, power of the BLRT and LMR tests was between .90 – 1.0 for most of the models. Small power was found in the LCA 10-item model with a sample size of 200 and in the LGM model with sample sizes of 200 and 500. In the case of the LCA model, power of LMR was .62 for categorical variables and .67 for continuous variables and for BLRT .84 and .98, respectively. In the case of the GMM with sample sizes of 200, 500 or 1000, power for the LMR and BLRT tests was .32 and .12, .76 and .66, or .97 and .97, respectively. These results mean that the BLRT test has more power in the LCA models and the LMR more power in the LGMM. When taking the results of higher Type I error rate for the LMR test into account, these results suggest that the BLRT is the best behaved test.

Fourth, a comparison of information criteria revealed that BIC behaves most consistently when deciding the number of latent classes (comparing 2 - 6 class solutions). When using AIC and adjusted BIC, conclusions tend to be biased to a larger number of latent classes. When comparing CAIC with BIC they behaved equally well for most of the models, except for the LGM model. In this case, BIC concluded in 6, 44 and 99 percent, and CAIC in 1, 22 and 97 percent of samples, true number of latent classes with sample sizes of 200, 500 or 1000, respectively.

Fifth, comparing the best information criteria BIC with the LMR and BLRT tests, the results were not so obvious: there appeared to be more than one significant result when comparing across different number of classes. Using the rule that the first non-significant result reveals the number of latent classes, the comparison revealed some differences. First, in the 8-item model, the BIC index behaved slightly better than the BLRT and this, in turn, behaved better than the LMR, especially when sample sizes were 200 or 500. In the 15-item model, all three compared indicators produced a high proportion (over .90) of true number of classes, when for BIC, this proportion was highest .99 – 1.0. For the 10-item categorical model, the BLRT behaved clearly better than the LMR test or BIC index when sample sizes were 200 or 500. In these cases, BIC produced .8 and .76 proportions for the right number of classes, LMR produced .43 and .72 proportions and BLRT produced .78 and .94. For the LGM model, all three indicators produced a low proportion of right number of classes. When a sample size was 200, 500 or 1000, BIC produced .06, .44 or .99 proportions, LMR produced .22, .63 or .73 proportions and VLMR produced .10, .58 or .87 proportions, respectively.

To conclude, the results of this second simulation study, the BIC index seems to behave well compared to other information criteria and is slightly better than BLRT or LMR for some models, whereas BLRT seems to behave best on average. For the LGM model, all the tests used did not produce satisfying results. Consequently, these results need to be more carefully examined in further simulation studies.

## **5. Monte Carlo simulation study**

A statistical theory usually provides information concerning the asymptotic properties of estimation methods but not, however, information how the estimators or test statistics behave with small sample sizes. Monte Carlo research is typically used to complement this lack of theoretical knowledge and, therefore, this method is also used in this study, described in Section 5.1. Monte Carlo simulations make it possible to get information of factors that have effects on the estimation of defined statistical models. The effects of different factors on the estimation are examined from three different viewpoints:

- (1) problems in the estimation of model parameters (Section 5.4.),
- (2) increasing or decreasing the ability to decide the number of latent classes (Section 5.5.),
- (3) a successful parameter estimation which is evaluated using different criteria (Section 5.6.) .

### **5.1. Monte Carlo method**

In the Monte Carlo method, a statistical model with fixed parameter values and observations with known distribution is defined. Observations are randomly generated according to these predefined model parameters. The data generated in this way simulates the sample and the defined model parameters represent the parameters in the population. The data are generated by this way  $R$  times. The parameters are estimated for each generated data by using a decided statistical model with a chosen estimation method. Information from each  $R$  replication, for example, information concerning the parameter values and standard errors, are then gathered together and averaged across the samples to get information of the bias and standard error of the estimation (Muthén & Muthén, 2002).

Designing a Monte Carlo study, there are several decisions to be made. In the following, the suggested steps to carry out a Monte Carlo study are listed. These steps are important when aiming to find practically meaningful and essential information about the factors related to modeling. According to Paxton, Curran, Bollen, Kirby, and Chen (2001) these steps are:

- 1) Setting a theoretically limited and well-grounded research question
- 2) Setting the model under investigation to represent as accurately as possible the model often found in practice
- 3) Defining specific research questions, such as sample sizes used and possible model misspecification
- 4) Defining parameter values in population
- 5) Choosing the suitable program
- 6) Carrying out a simulation study
- 7) Saving results
- 8) Finding out the possible problems and checking out the results
- 9) Presenting the results.

The following sections, 5.2 - 5.6, present an implementation of a Monte Carlo simulation study according to the steps presented above. The research questions and defined models are presented in section 5.2 respecting the first four steps. Then, the implementation of the simulation study is presented in section 5.3 respecting steps 5-7. When presenting results and answering to the research questions, the simulation study concentrates on three important aspects in the estimation of the LGM model a) problems in the parameter estimation, b) deciding the number of latent classes, and c) criteria used to evaluate the parameter estimation. These aspects are presented in sections 5.4 - 5.6. respecting steps 8 - 9.

## **5.2. Research questions and model under investigation**

The research aim of this study is to examine the functionality of the LGM model with a limited sample size, in which case asymptotic results are not in use. In practice, the LGM model usually has few measurement points. Consequently, the basic model of the present simulation study is limited to the linear LGM with four measurement points and two latent classes

$$\begin{cases} y_{it}^{(k)} = \eta_{0i}^{(k)} + (t-1)\eta_{1i}^{(k)} + \varepsilon_{it}^{(k)}, t = 1,2,3,4 \\ \eta_{0i}^{(k)} = \alpha_0^{(k)} + \zeta_{0i}^{(k)}, \\ \eta_{1i}^{(k)} = \alpha_1^{(k)} + \zeta_{1i}^{(k)}, k = 1,2; i = 1,2,\dots,n \end{cases} \quad (5.1)$$

in which  $y_{it}^{(k)}$  is  $i$ 's observation at time  $t$  from population  $k$ . This linear LG model (see Rovine & Molenaar, 1998) is often used in empirical studies (see for example Anstey, Hofer, Luszcz, 2003; Aunola, Leskinen, Lerkkanen & Nurmi, 2004; Colder, Mehta, Balanda, Campbell, Mayhew, Stanton, Pentz & Flay, 2001; Fuzhong, Barrera, Hops, Fisher, 2002; Duncan, Duncan, Alpert, Hops, Stoolmiller & Muthén, 1997; Li, Barrera, Hops, & Fisher, 2002; Muthén, Khoo, Francis, & Boscardin, 2003; Parrila, Aunola, Leskinen, Nurmi & Kirby, 2005).

In the present simulation study, the effects of the following five factors (a- e) on estimation are of interest and their values are, therefore, varied

- a) sample size is 50, 100, 200, 500 or 1000,
- b) the difference between expectations of latent components measured as SMD is 0.5, 1, 2, 3, 4 or 5, where SMD is the square root of Mahalanobis distance (McLahlan, 1999). SMD for latent components is

$$SMD = SMD(\boldsymbol{\eta}) = \sqrt{(\boldsymbol{\mu}_{\eta}^{(1)} - \boldsymbol{\mu}_{\eta}^{(2)})^T \boldsymbol{\Psi}^{-1} (\boldsymbol{\mu}_{\eta}^{(1)} - \boldsymbol{\mu}_{\eta}^{(2)})},$$

where

$\boldsymbol{\mu}_{\eta}^{(1)}$  and  $\boldsymbol{\mu}_{\eta}^{(2)}$  are expectation vectors of the latent components  $\boldsymbol{\eta}$  for classes 1 and 2 and the covariance matrix  $\boldsymbol{\Psi}$  of  $\boldsymbol{\eta}$  is equal in both latent classes,

- c) reliability of observed variables
 
$$rel(y_t) = 1 - \text{var}(\varepsilon_t) / \text{var}(y_t), t = 1, 2, 3, 4$$
 is low 0.5 or high 0.8,
- d) correlation between latent intercept and slope components is zero or 0.50,
- e) the number of measurements are four or seven,
- f) the proportion of class sizes are 1/3 and 2/3.

The effects on the estimation are examined in three ways:

- 1) the effect on problems appearing in estimation (see section 5.4),
- 2) the effect on deciding the number of latent classes (see section 5.5),
- 3) the effect on criteria used to evaluate parameter estimation (see section 5.6).

To examine the effect of these five factors described above, six different variations of the LGM model in equation 5.1 are used. In all models, sample size and SMD are varied according to a) and b). In the models named A.8 and A.5, the differences between two latent classes are in the mean of intercept component ( $\alpha_0^{(k)}$ ), and in the models named B.8 and B.5, also in the mean of slope component ( $\alpha_1^{(k)}$ ). In models A.8, A.5, B.8 and B.5, the correlation between the intercept and slope components is zero. The number in the model name indicates reliability of observed variables, which is high (.80) or low (.50). Model C.8 is similar to model A.8, except that the correlation between the intercept and slope is 0.50. Model A.5\* differs from the model A.5 in that this model has three additional measurement points locate in the middle of the four measurement times in A.5.

Table 5.1 presents the square root of the Mahalanobis distance for the observed and latent components. The square root of the Mahalanobis distance is also defined for the observed variables  $\mathbf{y}$  in the same way as for the latent components and is

$$SMD(\mathbf{y}) = \sqrt{(\boldsymbol{\mu}_y^{(1)} - \boldsymbol{\mu}_y^{(2)})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_y^{(1)} - \boldsymbol{\mu}_y^{(2)})},$$

where

$\boldsymbol{\mu}_y^{(1)}$  and  $\boldsymbol{\mu}_y^{(2)}$  are expectation vectors of the observed variables  $\mathbf{y}$ , for latent classes 1 and 2 and

$\boldsymbol{\Sigma}$  is covariance matrix of  $\mathbf{y}$  which is equal in both latent classes.

To refer to the distance of latent components  $SMD(\boldsymbol{\eta})$ , the term  $SMD$  is used, whereas when referring to the distance of observed variables, the term  $SMD(\mathbf{y})$  is used. In the following Table 5.1, the square root of the Mahalanobis distance of observed variables are presented for all models A.8 – C.8 with different  $SMD$ .

Table 5.1. Square root of the Mahalanobis distance of observed variables  $SMD(y)$  as a function of  $SMD$ .

Model	$SMD$				
	1	2	3	4	5
A.8	0.93	1.85	2.78	3.71	4.64
A.5	0.81	1.62	2.44	3.25	4.06
B.8	0.93	1.86	2.71	3.56	4.40
B.5	0.81	1.63	2.30	2.96	3.62
C.8	0.82	1.61	2.42	3.22	4.03
A.5*	0.86	1.72	2.58	3.44	4.30

The expected values of the latent and observed variables respecting  $SMD$  are presented in Table 5.2. For model A.5\*, the expected values are equal with the expected values for model A.5 in the time points 1, 2, 3 and 4.

Table 5.2. Expected values of latent components ( $\alpha$ ) and observed variables ( $\mu$ ) for model A.8, A.5, B.8, B.5 and C.8 with a five-set of parameter values used in the two latent classes.

Model A.8 and A.5					Time							
					1		2		3		4	
$SMD$	$\alpha_0^{(1)}$	$\alpha_1^{(1)}$	$\alpha_0^{(2)}$	$\alpha_1^{(2)}$	$\mu_1^{(1)}$	$\mu_1^{(2)}$	$\mu_2^{(1)}$	$\mu_2^{(2)}$	$\mu_3^{(1)}$	$\mu_3^{(2)}$	$\mu_4^{(1)}$	$\mu_4^{(2)}$
1	0	.2	1	.2	0	1	.2	1.2	.4	1.4	.6	1.6
2	0	.2	2	.2	0	2	.2	2.2	.4	2.4	.6	2.6
3	0	.2	3	.2	0	3	.2	3.2	.4	3.4	.6	3.6
4	0	.2	4	.2	0	4	.2	4.2	.4	4.4	.6	4.6
5	0	.2	5	.2	0	5	.2	5.2	.4	5.4	.6	5.6

Model B.8 and B.5					Time							
					1		2		3		4	
$SMD$	$\alpha_0^{(1)}$	$\alpha_1^{(1)}$	$\alpha_0^{(2)}$	$\alpha_1^{(2)}$	$\mu_1^{(1)}$	$\mu_1^{(2)}$	$\mu_2^{(1)}$	$\mu_2^{(2)}$	$\mu_3^{(1)}$	$\mu_3^{(2)}$	$\mu_4^{(1)}$	$\mu_4^{(2)}$
1	0	.2	1	0.200	0	1	.2	1.200	.4	1.400	.6	1.600
2	0	.2	1	0.975	0	1	.2	1.975	.4	2.950	.6	3.925
3	0	.2	1	1.465	0	1	.2	2.465	.4	3.930	.6	5.395
4	0	.2	1	1.932	0	1	.2	2.932	.4	4.864	.6	6.796
5	0	.2	1	2.391	0	1	.2	3.391	.4	5.782	.6	8.173

table continues



table continues

<b>Model C.8</b>					<b>Time</b>							
					<b>1</b>		<b>2</b>		<b>3</b>		<b>4</b>	
<i>SMD</i>	$\alpha_0^{(1)}$	$\alpha_1^{(1)}$	$\alpha_0^{(2)}$	$\alpha_1^{(2)}$	$\mu_1^{(1)}$	$\mu_1^{(2)}$	$\mu_2^{(1)}$	$\mu_2^{(2)}$	$\mu_3^{(1)}$	$\mu_3^{(2)}$	$\mu_4^{(1)}$	$\mu_4^{(2)}$
<b>1</b>	0	.2	0.866	.2	0	0.866	.2	1.066	.4	1.266	.6	1.466
<b>2</b>	0	.2	1.732	.2	0	1.732	.2	1.932	.4	2.132	.6	2.332
<b>3</b>	0	.2	2.598	.2	0	2.598	.2	2.798	.4	2.998	.6	3.198
<b>4</b>	0	.2	3.464	.2	0	3.464	.2	3.664	.4	3.864	.6	4.064
<b>5</b>	0	.2	4.330	.2	0	4.330	.2	4.530	.4	4.730	.6	4.930

By using presented values of  $\alpha_0^{(1)}, \alpha_0^{(2)}, \alpha_1^{(1)}, \alpha_1^{(2)}$  (Table 5.2), the models A.8, A.5, A.5\*, B.8, B.5 and C.8 are comparable in terms of square root of the Mahalanobis distance measured with latent components.

Theoretical covariance matrices of latent variables in the model A.8, A.5, A.5\*, B.8, B.5 and C.8 are same for both latent classes

$$\text{cov}(\mathbf{y}) = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}^T + \mathbf{\Theta},$$

where

$$\text{cov}(\boldsymbol{\eta}) = \mathbf{\Psi} = \begin{bmatrix} 1 & \\ 0 & .2 \end{bmatrix}$$

for models A.8, A.5, A.5\*, B.8 and B.5,

$$\text{cov}(\boldsymbol{\eta}) = \mathbf{\Psi} = \begin{bmatrix} 1 & \\ .224 & .2 \end{bmatrix}$$

for model C.8,

$$\mathbf{\Theta} = \text{diag}(.25 \ .30 \ .45 \ .70)$$

for models A.8 and B.8,

$$\Theta = \text{diag}(1.00 \ 1.20 \ 1.80 \ 2.80)$$

for models A.5 and B.5,

$$\Theta = \text{diag}(1.00 \ 1.05 \ 1.20 \ 1.45 \ 1.80 \ 2.25 \ 2.80)$$

for model A.5\* and

$$\Theta = \text{diag}(.25 \ .412 \ .674 \ 1.036)$$

for model C.8.

Then the values of the covariance matrices for the models A.8 and B.8 in both latent classes are

$$\text{cov}(\mathbf{y}) = \Sigma = \begin{bmatrix} 1.25 & & & \\ 1.00 & 1.50 & & \\ 1.00 & 1.40 & 2.25 & \\ 1.00 & 1.60 & 2.20 & 3.50 \end{bmatrix}.$$

The values of the covariance matrices for the models A.5 and B.5 in both latent classes are

$$\text{cov}(\mathbf{y}) = \Sigma = \begin{bmatrix} 2.00 & & & \\ 1.00 & 2.40 & & \\ 1.00 & 1.40 & 3.60 & \\ 1.00 & 1.60 & 2.20 & 5.60 \end{bmatrix}.$$

For model A.5\*, the covariance matrix is

$$\text{cov}(\mathbf{y}) = \Sigma = \begin{bmatrix} 2.00 & & & & & & & \\ 1.00 & 2.10 & & & & & & \\ 1.00 & 1.10 & 2.40 & & & & & \\ 1.00 & 1.15 & 1.30 & 2.90 & & & & \\ 1.00 & 1.20 & 1.40 & 1.60 & 3.60 & & & \\ 1.00 & 1.25 & 1.50 & 1.75 & 2.00 & 4.50 & & \\ 1.00 & 1.30 & 1.60 & 1.90 & 2.20 & 2.50 & 5.60 & \end{bmatrix}$$

and for model C.8 the covariance matrix is

$$\text{cov}(\mathbf{y}) = \Sigma = \begin{bmatrix} 1.250 & & & & \\ 1.224 & 2.060 & & & \\ 1.448 & 2.072 & 3.370 & & \\ 1.672 & 2.496 & 3.320 & 5.180 & \end{bmatrix}.$$

The LGM models above are the models used in the simulation study of this research. For simulated data, either a true two-latent-class model using true starting values or, alternatively, a wrong three-latent-class model using random starting values is estimated. Sample sizes of the simulation study for each model are  $n = 50, 100, 200, 500$  or  $1000$ .

### 5.3. Implementation of the simulation study

The simulated data are produced and analysed by using the Mplus program (Muthén & Muthén, 1998-2006). The first simulation is carried out using version 3.01 and the latest simulation using version 4.2. Some simulations, which were in the first place carried out using the earlier version, were repeated afterwards using version 4.2 to ensure that results are non-changeable. These results are presented later in the section.

For each defined model, A.8, A.5, A.5\*, B.8, B.5 and C.8, random data are generated using the Mplus program. This generated data are then estimated using the MLR estimator and the results of the parameter estimates, standard errors of the parameter estimates and the log-likelihoods are saved. This process is replicated 10000 times and information of replications is saved successively to two files. The data generation and estimation is defined with one script. An example of this type of script is presented in Appendix 1.

The example run (Appendix 1) generates and estimates data according to the parameter values of model A.8 when SMD is 3 and the sample size is 500. The run produces two files in which the results of 10000 replications are saved. The first output file has tabled information for criteria to evaluate parameter estimation. This file also has information related to the number of latent classes from each replication. This text file is cleaned with written macro using text editor and analysed after that using the SPSS program (see Appendix 2).

The other file produced by script (see Appendix 1) includes replication number, information concerning convergence, the value of log-likelihood, and the parameter estimates for each replication. This file is analyzed with the SPSS script shown in Appendix 3 by producing the number of negative variance estimates in all and for each parameter separately.

The results are then gathered to the tables, including information about the problems in estimation (see section 5.4), deciding the number of latent classes (see section 5.5) and the criteria used to evaluate the parameter estimation (see section 5.6).

## **5.4. Problems in estimation of latent growth mixture model**

A situation where an estimation of the model is not convergent is not unusual in empirical research with SEM. The reasons behind nonconvergence are usually poor data, a misspecified model, or poor starting values of the parameters. For example, if the model includes very weak associations, or if the variances of observed variables in the model are at totally different scales, the role of the starting values on convergence is particularly important.

The model estimation can also produce parameter estimates which are not admissible, that is, the estimated correlation may be larger than one, or the estimated variance can be negative. The appearance of negative variance estimates

may also be due to normal variations in sampling, the model's unidentification, or outliers (Chen, Bollen, Paxton, Curran & Kirby, 2001). The problem of negative variance estimates, in particular, has been under discussion. This discussion can be summarized into the next three questions (Chen et al., 2001):

- a) when are negative variance estimates most likely to appear?
- b) what are their consequences?
- c) what would be the most appropriate way to proceed with negative variance estimates?

According to Chen et al. (2001), the negative variance estimates have been shown to appear typically with small sample sizes, which can be related to the variation of sampling. The probability of negative variance estimates was not directly related to the degree of misspecified model, but rather the misspecified model influenced the values of parameter estimates, their standard errors and the distribution of error variances. The conclusion is that negative variances should not be interpreted as a misspecified model. On the other hand, acceptable values do not automatically tell that the model is properly specified either.

According to Chen et al. (2001), the challenge of future research is; how the overparametrized model, either defined by extra paths or covariance structure of error, impact on the success of the model estimation. On the basis of research (Tolvanen, 2000) it seems that in the context of LGM, the overparametrized model produces, on average, more negative variance estimates than a true model. This is also due to growing standard errors of parameter estimates.

## **5.5. Deciding on the number of latent classes**

In mixture analysis, the evaluation of the overall fit of the estimated model is not possible in the same way that it usually is in the SEM framework. Also, the use of the likelihood ratio test to evaluate the right number of latent classes, is not appropriate in mixture analyses. Instead, to evaluate the model fit and the right number of latent classes in mixture analysis, the following three information criteria and three statistical tests are used (see section 3.2.2):

- 1) Akaike's information criteria (AIC) (Akaike, 1987),
- 2) Bayes information criteria (BIC) (Schwartz, 1978),
- 3) adjusted BIC (aBIC) (Sclove, 1987),
- 4) Vuong, Lo, Mendell & Rubin (VLMR) (Vuong, 1989) test,

- 5) adjusted VLMR (LMR) (Lo, Mendell and Rubin, 2001) test and
- 6) parametric bootstrapped likelihood ratio test (BLRT) (McLahlan & Peel, 2000; Muthén & Muthén, 1998-2006).

In this study, ending up to a certain number of latent classes is presented as a proportion of replications for all of the above mentioned six possibilities. The failed estimations are out of calculation of this proportion, but the negative variance estimates are included, because they are seen to represent normal variation in sampling (Chen et. al., 2001).

In this study, there are three questions related to deciding the right number of latent classes, namely:

- a) what is the proportion of replications ending up to the right two-class solution when the data have in truth two-latent classes?
- b) what is the proportion of replications ending up to wrong three number of classes when the data have in truth two-latent classes?

The answers to these two questions will give information about what sample size is the needed sample size to achieve .70, .80 or even .90 power in order to detect the right number of latent classes using the three information criteria 1-3 or statistical tests 4-6, and which of them are most useful when deciding the number of latent classes in a linear LGMM.

## 5.6. Criteria used to evaluate parameter estimation

The following four criteria (c1, c2, c3 and c4) are generally used to evaluate the goodness and the validity of the LGMM estimation and the properties of MLR estimator.

Denote freely estimated parameters as a  $P \times 1$  vector

$$\boldsymbol{\tau} = \text{vec}(\boldsymbol{\alpha}, \boldsymbol{\Psi}, \boldsymbol{\Theta}).$$

First criterion (c1) is the MSE (mean square error) of the estimator  $\hat{\tau}_p$

$$MSE(\hat{\tau}_p) = E(\hat{\tau}_p - \tau_p)^2, \quad p = 1, 2, \dots, P.$$

It is well known that

$$MSE(\hat{\tau}_p) = \text{var}(\hat{\tau}_p) + B^2(\hat{\tau}_p),$$

in which bias of parameter estimator is

$$B(\hat{\tau}_p) = E(\hat{\tau}_p) - \tau_p.$$

Estimates of MSE and its components are defined as follows

$$\hat{MSE}(\hat{\tau}_p) = \frac{1}{R} \sum_{r=1}^R (\hat{\tau}_{pr} - \tau_p)^2 = \hat{\text{var}}(\hat{\tau}_p) + \hat{B}^2(\hat{\tau}_p), \quad (c1)$$

where

$$\hat{\text{var}}(\hat{\tau}_p) = \frac{1}{R} \sum_{r=1}^R (\hat{\tau}_{pr} - E(\hat{\tau}_p))^2,$$

$$\hat{B}(\hat{\tau}_p) = E(\hat{\tau}_p) - \tau_p$$

and

$$E(\hat{\tau}_p) = \frac{1}{R} \sum_{r=1}^R \hat{\tau}_p^{(r)}.$$

Second criterion (c2) is the proportion of bias in the MSE

$$B^2(\hat{\tau}_p) / MSE(\hat{\tau}_p), p = 1, 2, \dots, P.$$

The proportion of bias in the MSE is estimated as

$$PB = \hat{B}^2(\hat{\tau}_p) / \hat{MSE}(\hat{\tau}_p), p = 1, 2, \dots, P. \quad (c2)$$

Third estimation criterion (c3) is the relative bias of asymptotic standard error of parameter estimates. This relative bias is calculated by dividing the expected asymptotic standard error of the parameter estimator by the standard deviation of the parameter estimator

$$RB\left(\widehat{ase}(\widehat{\tau}_p)\right) = \frac{E\left(ase(\widehat{\tau}_{pr})\right)}{SD(\widehat{\tau}_p)} .$$

This relative bias is estimated by dividing the average of estimated standard errors across the valid replications, with the estimated standard deviation of the parameter estimate calculated across the valid replications

$$\widehat{RB}\left(\widehat{ase}(\widehat{\tau}_p)\right) = \frac{\frac{1}{R} \sum_{r=1}^R \widehat{ase}(\widehat{\tau}_{pr})}{\widehat{SD}(\widehat{\tau}_p)} . \quad (c3)$$

Because the number of replications is large, the estimated standard deviation  $\widehat{SD}(\widehat{\tau}_p)$  can accurately approximate the true standard deviation. Therefore, this

criterion  $\widehat{RB}$  describes the bias of standard error of parameter estimator as its proportion to true standard error. If this proportion is lower than one, the standard error is biased downward and produces smaller confidence intervals for parameter estimates. If this proportion is greater than one, the standard error is biased upward and produces larger confidence intervals for parameter estimates. If standard error estimate is violated, this also has consequences to the testing of the true value of parameter estimates. When the standard error estimate is biased downward and  $RB = .95$ , this leads to the change of p-value from nominal .05 level to .063 level, from nominal .01 level to .014 level, and from nominal .001 level to .0018 level. When the standard error estimate is biased downward and  $RB = .90$  percent from its true value, this leads to the change of p-value from nominal .05 level to .078 level, from nominal .01 level to .020 level, and from nominal .001 level to .003 level.

The fourth criterion is the 95 % coverage, which gives the proportion of replications in which the true value of parameter falls into the estimated 95 % confidence interval

$$coverage(\widehat{\tau}_p) = \frac{1}{R} \sum_{r=1}^R CI_{95}(\widehat{\tau}_{pr}) , \quad (c4)$$

where

$$CI_{95}(\widehat{\tau}_{pr}) = 1 ,$$



if

$$\tau_p - 1.96 \times \widehat{ase}(\hat{\tau}_{pr}) < \hat{\tau}_{pr} < \tau_p + 1.96 \times \widehat{ase}(\hat{\tau}_{pr})$$

and

$$CI_{95}(\hat{\tau}_{pr}) = 0 \text{ elsewhere,}$$

$$p = 1, 2, \dots, P, \quad r = 1, 2, \dots, R.$$

## 6. Results of the simulation study

Before starting a simulation study, one should decide the complexity of the examined models. In the present study, this complexity is first examined by a pilot simulation study, using model A.8 with three different equality constraints between the latent classes. After this a broader simulation study, including other models, is carried out.

To get a more comprehensive view of the factors affecting latent growth mixture modeling, the results of basic model A.8 are first described. In this context, the impact of sample size ( $n = 50, 100, 200, 500$  or  $1000$ ), on one hand, and the impact of distance between two groups ( $SMD = 1, 2, 3, 4$  and  $5$ ), on the other, on the success and validity of estimation are considered. The success and validity of LGMM is examined in terms of the number of failed estimations and improper solutions (section 6.1), power to find the right number of latent classes (section 6.2), and the evaluation of parameter estimation (section 6.3).

In addition to sample size and SMD, the effect of reliability of observed variables is examined, changing the .80 reliability of observed variables in model A.8 to .50 in model A.5. Moreover, the effect of additional measurement points is examined using additional measurement points in model A.5\*. These effects are expressed in their own paragraphs in sections 6.1, 6.2 and 6.3.

Finally, the effect of model construct is examined comparing models C.8 vs. A.8, B.8 vs. A.8, and B.5 vs. A.5. When in models A.8 and A.5 latent groups differ according to the mean of intercept component ( $\alpha_0$ ), in models B.8 and B.5 groups differ from each others also according to the mean of slope component ( $\alpha_1$ ). Further, model C.8, where latent components correlate with each other, is compared with model A.8, where correlation of latent components is zero. These effects are also described in their own paragraphs in sections 6.1, 6.2 and 6.3.

## 6.1. Problems in estimation of latent growth mixture model

Problems in the estimation are considered from two viewpoints: the number of failed estimations and the number of negative variance estimates. These two problems are first mapped with a pilot simulation study. The goal of the pilot simulation study, using model A.8, is therefore twofold: On one hand, the aim is to find the boundaries for true models, where estimation proceeds successfully. On the other hand, the aim is to find boundaries where the estimated parameter values of the models are acceptable (e.g., estimated variances of the models used are positive in most cases of the samples). The results of this pilot simulation study are presented in section 6.1.1. Based on the pilot simulation study, the investigation of the problems in estimation is extended to all other models. These results are presented in section 6.1.2.

### 6.1.1. Pilot simulation study

The pilot simulation study uses model A.8, in which SMD varies between 0.5 to 5 depending on the value of  $\alpha_0^{(2)}$ . Sample sizes of generated data are  $n = 50, 100, 200, 500$  or  $1000$ , in which the observations are received from a random indicator for two classes. This random indicator is binomially distributed  $\text{Bin}(n, 1/3)$ . Each of the generated data is replicated 10000 times.

LGM model A.8 is estimated for a two group solution in three situations I-III:

- I) all parameters  $\alpha_0, \alpha_1, \psi_{11}, \psi_{12}, \psi_{22}, \theta_1, \theta_2, \theta_3$  and  $\theta_4$  are freely estimated in both latent classes
- II) the parameters  $\alpha_0, \alpha_1, \psi_{11}, \psi_{12}$  and  $\psi_{22}$  are freely estimated in both latent classes and parameters  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  are fixed equal between latent classes
- III) the parameters  $\alpha_0$  and  $\alpha_1$  are freely estimated in both latent classes and the parameters  $\psi_{11}, \psi_{12}, \psi_{22}, \theta_1, \theta_2, \theta_3$  and  $\theta_4$  are fixed equal between latent classes.

Because the starting values of the parameters for latent classes are true population values and the estimated model has a real number of groups, the reason why the convergence of estimation fails may be an empirical non-identification of the

model. This, in turn, is probably due to the fact that the proportion of one of the classes is too small. In Table 6.1., the estimation results of 10000 samples are presented with different SMDs and with varied sample sizes.

Because simulations are carried out with two different versions of Mplus program (i.e., earlier version 3.01 and current version 4.2.), the results concerning failed estimation and negative variance estimates in Table 6.1., are presented in two columns. As can be seen from Table 6.1., the number of negative variances appears frequently in situations I and II, whereas in situation III they are frequent only when the sample size is small.

Table 6.1. Percentages of failed estimations (FE) and negative variance estimates (NV) with varied sample size and SMD. Model A.8, R=10000.

n	SMD	Situation I				Situation II				Situation III			
		FE	FE*	NV	NV*	FE	FE*	NV	NV*	FE	FE*	NV	NV*
50	0.5	<b>2.38</b>	<b>2.43</b>	<b>78.98</b>	<b>78.93</b>	<b>5.77</b>	<b>5.94</b>	<b>37.48</b>	<b>37.35</b>	0.08	0.07	6.53	6.55
100	0.5	<b>1.96</b>	<b>2.00</b>	<b>63.73</b>	<b>63.69</b>	<b>4.44</b>	<b>4.44</b>	<b>26.10</b>	<b>26.21</b>	0.19	0.19	0.75	0.71
200	0.5	<b>1.79</b>	<b>1.79</b>	<b>52.80</b>	<b>52.89</b>	<b>4.16</b>	<b>3.69</b>	<b>20.98</b>	<b>21.33</b>	0.36	0.31	0.03	0.03
500	0.5	<b>1.44</b>	<b>1.54</b>	<b>37.56</b>	<b>45.38</b>	<b>2.64</b>	<b>2.70</b>	<b>18.62</b>	<b>18.62</b>	0.69	0.67	0	0
1000	0.5	<b>1.25</b>	<b>1.17</b>	<b>41.79</b>	<b>42.15</b>	<b>2.14</b>	<b>2.19</b>	<b>17.36</b>	<b>17.45</b>	0.87	0.89	0	0
50	1	<b>2.39</b>	<b>2.36</b>	<b>76.17</b>	<b>76.24</b>	<b>5.48</b>	<b>5.30</b>	<b>34.94</b>	<b>34.95</b>	0.04	0.05	6.95	6.94
100	1	<b>2.13</b>	<b>1.97</b>	<b>60.50</b>	<b>60.66</b>	<b>3.88</b>	<b>3.90</b>	<b>24.52</b>	<b>26.45</b>	0.16	0.10	0.88	0.88
200	1	<b>1.63</b>	<b>1.64</b>	<b>49.99</b>	<b>49.98</b>	<b>3.49</b>	<b>3.53</b>	<b>19.07</b>	<b>19.03</b>	0.20	0.20	0.03	0.03
500	1	<b>1.24</b>	<b>1.29</b>	<b>42.66</b>	<b>42.63</b>	<b>2.01</b>	<b>2.08</b>	<b>17.25</b>	<b>17.24</b>	0.30	0.26	0	0
1000	1	<b>1.05</b>	<b>1.06</b>	<b>38.52</b>	<b>38.63</b>	<b>1.49</b>	<b>1.53</b>	<b>15.78</b>	<b>15.75</b>	0.23	0.19	0	0
50	2	<b>2.29</b>	<b>2.32</b>	<b>69.18</b>	<b>69.21</b>	<b>3.98</b>	<b>4.04</b>	<b>29.36</b>	<b>29.51</b>	0	0.01	7.26	7.26
100	2	<b>1.36</b>	<b>1.41</b>	<b>51.26</b>	<b>51.23</b>	<b>2.66</b>	<b>2.72</b>	<b>16.97</b>	<b>16.95</b>	0.01	0.01	1.04	1.04
200	2	<b>0.75</b>	<b>0.83</b>	<b>36.11</b>	<b>36.07</b>	<b>0.96</b>	<b>1.06</b>	<b>10.98</b>	<b>10.91</b>	0	0.01	0.02	0.02
500	2	<b>0.24</b>	<b>0.22</b>	<b>17.67</b>	<b>17.70</b>	0.16	0.15	4.13	4.14	0	0.0	0	0.0
1000	2	0.05	0.02	6.58	6.58	0.01	0.0	1.19	1.20	0	0	0	0
50	3	<b>0.71</b>	<b>0.83</b>	<b>58.50</b>	<b>58.38</b>	<b>0.20</b>	<b>1.22</b>	<b>20.35</b>	<b>20.36</b>	0	0	6.39	6.39
100	3	<b>0.19</b>	<b>0.18</b>	<b>30.63</b>	<b>30.62</b>	0.16	0.18	5.70	5.71	0.01	0	1.10	1.10
200	3	0	0.02	9.14	9.13	0.01	0.01	0.82	0.82	0	0	0.03	0.03
500	3	0	0	0.34	0.34	0	0	0.01	0.01	0	0	0	0
1000	3	0	0.01	0.01	0.01	0	0	0	0	0	0	0	0
50	4	<b>0.05</b>	<b>0.07</b>	<b>44.85</b>	<b>44.83</b>	<b>0.04</b>	<b>0.04</b>	<b>11.24</b>	<b>11.25</b>	0	0	5.17	5.17
100	4	<b>0</b>	<b>0</b>	<b>15.52</b>	<b>15.52</b>	0	0	1.33	1.33	0	0	0.81	0.81
200	4	0	0	3.16	3.16	0	0	0.02	0.02	0	0	0.03	0.03
500	4	0	0	0.01	0.01	0	0	0	0	0	0	0	0
1000	4	0	0	0	0	0	0	0	0	0	0	0	0
50	5	<b>0</b>	<b>0</b>	<b>37.48</b>	<b>37.48</b>	0.01	0.01	7.59	7.59	0	0	4.34	4.34
100	5	<b>0</b>	<b>0</b>	<b>12.17</b>	<b>12.17</b>	0	0	0.75	0.75	0	0	0.59	0.59
200	5	0	0	2.15	2.15	0	0	0.01	0.01	0	0	0.02	0.02
500	5	0	0	0.04	0.04	0	0	0	0	0	0	0	0
1000	5	0	0	0	0	0	0	0	0	0	0	0	0

\* Estimated with Mplus version 4.2. Note. The values in which the number of negative variances exceed 10 % are bolded.

When the distance between the classes diminishes, the convergence of estimation fails more often in situation II, than in other situations. In situation I and II, the number of failed estimation decreases when the sample size increases, whereas in situation III the number of failed estimation slightly unexpectedly increases. If the difference between the groups is small ( $SMD \leq 1$ ), the number of failed estimations varies depending on the sample size used, up to 5.9 percent of replications when using version 4.2, and up to 5.8 percent when using version 3.01. When the distance between classes is clear ( $SMD \geq 3$ ) and the estimated model has few parameters (Situation III), the number of failed estimations is lower than 0.01 percent of replications.

In the most complex situation (Situation I) the number of negative variances is high. In this case, the number of negative variances decreases when SMD and sample size increase. Despite the high distance between the classes ( $SMD = 5$ ), the number of negative variances is over 12 % when the sample size is 100.

When error variances are fixed to be equal between the classes in situation II, the number of negative variance estimates clearly decreases when compared with situation I. In this case the number of negative variance estimates is over 10 %, if the distance between the groups is small ( $SMD = 0.5$  or  $1$ ). If  $SMD = 2$  and the number of cases is at minimum 500, the proportion of negative variance estimates is less than 5 %. If the difference between the groups is clear ( $SMD \geq 3$ ), the proportion of negative variance estimates is lower than 6 % when the sample size is 100 at minimum.

If the model consists of fewer free parameters (Situation III), the number of unacceptable models varies from 0% to 7.3 % depending on SMD and sample size.

The results presented above reveal that the problems in LGMM estimation mainly appear in negative variance estimates. In situations I and II, this is evident particularly with small sample sizes. Because the present simulation study concentrates on the behavior of LGMM with small sample sizes, in the following simulations using models A.8, A.5, A.5\*, B.8, B.5 and C.8, the parameters are constrained to be equal between the latent classes, as in situation III. In situation III, the numbers of negative variance estimates are rare enough to make it possible to examine other criteria of estimation.

### **6.1.2. Problems in estimation of latent growth mixture model in the main simulation study**

In the main simulation study, the parameter estimates are fixed equal between the classes, as in situation III in the pilot simulation study. When  $SMD=1$ , the estimation fails for model A.8, A.5 and C.8, but the number of failed estimations is lower than 0.3, 0.5 or 0.4 percent of replications, respectively. In turn, negative variance estimates seem to appear frequently with small sample sizes. These results for models A.8, A.5, A.5\*, B.8, B.5 and C.8 are shown in Table 6.2. When  $SMD \geq 2$ , estimation fails only in few cases for all sample sizes and for all tested models A.8, A.5, A.5\*, B.8, B.5 and C.8.

Negative variance estimates in model A.8 appear in 4.3 – 7.3 percent of replications when  $n = 50$ , in 0.6 – 1.1 percent of replications when  $n = 100$ , and are rare when  $n \geq 200$ .

The effect of low reliability (A.8 vs. A.5) on the number of negative variance estimates is strong. When reliability decreases, the number of negative variance estimates increases. If  $SMD$  is small, negative variance estimates appears also with larger sample sizes.

With a small sample size  $n = 50$ , the proportion of negative variance estimates in model A.5 is 38.0 percent of replications, when  $SMD$  is 1 and decreases to 11.2 percent when  $SMD$  increases to 5. These proportions are 2.6 – 5.5 times larger in model A.5 than in model A.8. When  $n = 100$ , the proportion of negative variance estimates rapidly decreases from 20 to 2.2 percent, when  $SMD$  increases from 1 to 5. Unlike in model A.8, in model A.5, negative variance estimates appear also when  $n = 200, 500$  or  $1000$ , in which case the percentage of negative variance estimate is 8.7, 3.0 or 0.8 when  $SMD$  is 1, and 3.7, 0.5 or 0.03 when  $SMD$  is 2, respectively.

When adding additional measurement points (A.5\* vs. A.5), the number of negative variance estimates dramatically decreases. When  $SMD$  is 1 or 2, this decrease is 60 - 68 percent. When  $SMD \geq 3$ , the number of negative variance estimates is even lower than in model A.8 and decreases to 0.6 - 3.7 percent when  $n = 50$ .

Table 6.2. The number of negative variance estimates in model A.8, A.5, A.5\*, B.8, B.5 and C.8, based on varying numbers of replications.

n	SMD	A.8	A.5	A.5*	B.8	B.5	C.8
50	1	<b>695</b>	<b>3797</b>	<b>1427</b>	<b>695</b>	<b>3797</b>	<b>854</b>
100	1	88	<b>1995</b>	<b>622</b>	88	<b>1995</b>	147
200	1	3	870	171	3	870	7
500	1	0	295	8	0	295	0
1000	1	0	83	0	0	83	0
50	2	<b>726</b>	<b>2842</b>	<b>896</b>	<b>579</b>	<b>3344</b>	<b>894</b>
100	2	104	<b>1125</b>	309	53	<b>1720</b>	157
200	2	2	374	52	1	<b>683</b>	12
500	2	0	52	0	0	73	0
1000	2	0	3	0	0	2	0
50	3	<b>639</b>	<b>1941</b>	370	<b>502</b>	<b>2491</b>	<b>878</b>
100	3	110	484	32	53	<b>972</b>	177
200	3	3	44	0	1	194	9
500	3	0	0	0	0	3	0
1000	3	0	0	0	0	0	0
50	4	<b>517</b>	<b>1374</b>	115	453	<b>1751</b>	<b>722</b>
100	4	81	250	2	49	489	126
200	4	3	15	0	1	54	7
500	4	0	0	0	0	0	0
1000	4	0	0	0	0	0	0
50	5	434	<b>1119</b>	62	420	<b>1325</b>	<b>576</b>
100	5	59	219	2	48	302	99
200	5	2	13	0	1	31	6
500	5	0	0	0	0	0	0
1000	5	0	0	0	0	0	0

Note. If the amount of negative variance estimate is greater than 5 percent, the cell is bolded.

The number of negative variance estimates is lower in model B.8 than in model A.8. However, the number of negative variances is higher in model B.5 compared with model A.5. The differences between the models are largest when SMD is 2, 3 or 4 and n = 50 or 100. When n = 50, the number of negative variance estimates is 12 - 21 percent lower and, when n = 100, 40 – 52 percent lower in model B.8 than in model A.8. When n = 50, the number of negative variance estimates is 18 - 28

percent higher, and when  $n = 100$ , 38 – 101 percent higher in model B.5 than in model A.5.

The number of negative variance estimates is greater in model C.8 than in model A.8 and, as in model A.8, is quite a few also in model C.8 when  $n \geq 200$ . In model C.8, the number of negative variance estimates is 23 – 40 percent greater when  $n = 50$ , and 54 – 68 percent greater when  $n = 100$ , than in model A.8.

Negative variance estimates appear most often for  $\theta_1$ ,  $\psi_{00}$  and  $\psi_{11}$  parameters. However, when the sample size is small  $n = 50$ , negative variance estimates appear in 11.4 -17.3 percent of replications for  $\theta_4$  parameter in model B.8, whereas in other models appearance is lower than 0.9 percent of replications. If the number of negative variance estimates is greater than one percent of replications, the values in the Table 6.3 are written in bold and their distribution is discussed in the following three paragraphs.

Negative variance estimates mostly occur in error variance of the first measurement in model A.8, especially when  $SMD \geq 2$ . They are rare in variances of latent components and, when occurring, they are related to the small sample size  $n = 50$ , and a low difference between latent classes ( $SMD = 1$  or  $2$ ). In this case, negative variance estimates appear for variance of slope and are evident in 9.1 - 17.4 percent of the cases.

The effect of reliability is seen when comparing model A.8 with model A.5: in model A.5, negative variance estimates appear in the variances of latent factors rather than in error variances. This appearance is mostly seen in the slope component, where the results suggest to reject this component from the model, in practice.

As in model A.8, in model C.8, negative variance estimates occur mostly in error variance of the first measurement. In model C.8, the percentage of negative variances of  $\theta_1$  parameter is lower, and that of  $\psi_{11}$  parameter clearly greater, than in model A.8. As in model A.8, in model B.8, negative variance estimates occur mostly in error variance of the first measurement. In model B.8, the percentage of negative estimates of  $\theta_1$  parameter is lower, and that of  $\psi_{11}$  parameter greater, than in model A.8. As in model A.5, in model B.5, negative variance estimates occur mostly in error variance of latent components. In model B.5, the proportion of negative estimates of  $\psi_{00}$  parameter is lower, and that of  $\psi_{11}$  parameter greater, than in model A.5.



Table 6.3. The percentages of negative variance estimates for  $\theta_1, \psi_{00}$  and  $\psi_{11}$  parameters from the replications in which negative variance estimates appear (see Table 6.2).

Model		A.8			A.5			A.5*			B.8			B.5			C.8		
n	SMD	$\theta_1$	$\psi_{00}$	$\psi_{11}$	$\theta_1$	$\psi_{00}$	$\psi_{11}$	$\theta_1$	$\psi_{00}$	$\psi_{11}$	$\theta_1$	$\psi_{00}$	$\psi_{11}$	$\theta_1$	$\psi_{00}$	$\psi_{11}$	$\theta_1$	$\psi_{00}$	$\psi_{11}$
50	1	<b>75.3</b>	<b>3.6</b>	<b>17.4</b>	<b>3.2</b>	<b>55.1</b>	<b>58.0</b>	<b>0.0</b>	<b>7.3</b>	<b>92.5</b>	<b>75.3</b>	<b>3.6</b>	<b>17.4</b>	<b>3.2</b>	<b>55.1</b>	<b>58.0</b>	<b>63.8</b>	<b>7.1</b>	<b>27.9</b>
100	1	71.6	0.0	28.4	<b>0.2</b>	<b>36.5</b>	<b>69.9</b>	<b>0.0</b>	<b>0.5</b>	<b>99.5</b>	71.6	0.0	28.4	<b>0.2</b>	<b>36.5</b>	<b>69.9</b>	<b>47.6</b>	<b>0.0</b>	<b>52.4</b>
200	1	33.3	0.0	66.7	<b>0.0</b>	<b>17.2</b>	<b>84.1</b>	<b>0.0</b>	<b>0.0</b>	<b>100</b>	33.3	0.0	66.7	<b>0.0</b>	<b>17.2</b>	<b>84.1</b>	14.3	0.0	85.7
500	1	-	-	-	<b>0.0</b>	<b>0.7</b>	<b>99.3</b>	0.0	0.0	100	-	-	-	<b>0.0</b>	<b>0.7</b>	<b>99.3</b>	-	-	-
1000	1	-	-	-	0.0	0.0	100	-	-	-	-	-	-	0.0	0.0	100	-	-	-
50	2	<b>86.2</b>	<b>0.1</b>	<b>9.1</b>	<b>5.8</b>	<b>36.4</b>	<b>69.4</b>	<b>0.1</b>	<b>1.00</b>	<b>98.9</b>	<b>67.5</b>	<b>1.0</b>	<b>19.3</b>	<b>2.3</b>	<b>24.5</b>	<b>76.1</b>	<b>76.7</b>	<b>0.4</b>	<b>21.5</b>
100	2	<b>90.4</b>	<b>0.0</b>	<b>9.6</b>	<b>0.7</b>	<b>18.0</b>	<b>84.6</b>	<b>0.0</b>	<b>0.0</b>	<b>100</b>	86.8	0.0	9.4	<b>0.1</b>	<b>14.9</b>	<b>85.1</b>	<b>73.2</b>	<b>0.0</b>	<b>26.8</b>
200	2	100	0.0	0.0	<b>0.0</b>	<b>3.2</b>	<b>97.1</b>	0.0	0.0	100	100	0.0	0.0	<b>0.0</b>	<b>7.0</b>	<b>93.0</b>	41.7	0.0	58.3
500	2	-	-	-	0.0	0.0	100	-	-	-	-	-	-	0.0	0.0	100	-	-	-
1000	2	-	-	-	0.0	0.0	100	-	-	-	-	-	-	0.0	0.0	100	-	-	-
50	3	<b>93.6</b>	<b>0.0</b>	<b>2.2</b>	<b>9.7</b>	<b>23.9</b>	<b>74.2</b>	<b>0.0</b>	<b>0.0</b>	<b>100</b>	<b>79.1</b>	<b>0.4</b>	<b>3.0</b>	<b>2.8</b>	<b>19.0</b>	<b>79.7</b>	<b>89.0</b>	<b>0.0</b>	<b>10.1</b>
100	3	<b>99.1</b>	<b>0.0</b>	<b>0.9</b>	<b>2.1</b>	<b>8.9</b>	<b>90.9</b>	0.0	0.0	100	86.8	0.0	9.4	<b>0.2</b>	<b>9.8</b>	<b>90.1</b>	<b>93.8</b>	<b>0.0</b>	<b>6.2</b>
200	3	100	0.0	0.0	2.3	0.0	97.7	-	-	-	100	0.0	0.0	<b>0.0</b>	<b>4.6</b>	<b>95.4</b>	100	0.0	0.0
500	3	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	100	-	-	-
1000	3	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	0.0	-	-	-
50	4	<b>94.8</b>	<b>0.0</b>	<b>0.2</b>	<b>12.0</b>	<b>16.2</b>	<b>77.4</b>	<b>0.0</b>	<b>0.0</b>	<b>100</b>	<b>82.8</b>	<b>0.2</b>	<b>0.9</b>	<b>4.0</b>	<b>14.7</b>	<b>82.8</b>	<b>96.0</b>	<b>0.0</b>	<b>3.3</b>
100	4	100	0.0	0.0	<b>3.2</b>	<b>6.8</b>	<b>91.2</b>	0.0	0.0	100	98.0	0.0	0.0	<b>0.2</b>	<b>3.5</b>	<b>96.3</b>	<b>100</b>	<b>0.0</b>	<b>0.0</b>
200	4	100	0.0	0.0	0.0	0.0	100	-	-	-	100	0.0	0.0	0.0	0.0	100	100	0.0	0.0
500	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1000	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
50	5	<b>94.9</b>	<b>0.0</b>	<b>0.0</b>	<b>9.7</b>	<b>12.9</b>	<b>82.0</b>	0.0	0.0	100	<b>88.1</b>	<b>0.0</b>	<b>0.5</b>	<b>4.7</b>	<b>11.2</b>	<b>86.3</b>	<b>96.0</b>	<b>0.0</b>	<b>2.8</b>
100	5	100	0.0	0.0	<b>2.3</b>	<b>4.1</b>	<b>94.5</b>	0.0	0.0	100	97.9	0.0	0.0	<b>0.3</b>	<b>1.7</b>	<b>98.0</b>	100	0.0	0.0
200	5	100	0.0	0.0	0.0	0.0	100	-	-	-	100	0.0	0.0	0.0	0.0	100	100	0.0	0.0
500	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1000	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Note 1. If the number of negative variances is greater than one percent of 10000 replications, the values in the table are in bold. Note 2. The mark ‘-’ means that there are no negative variance estimates.

Because negative variances are expected to occur when the sample size is small, due to normal variation in sampling (Chen et al., 2001), they are included into the next analysis, which estimates the power of the VLMR, LMR and BLRT tests and analyzes the behavior of information criteria as well as the behavior of parameter

estimates. In the case of failed estimation, they do not provide information related to power of tests or information criteria and, therefore, are out of the use.

## **6.2. Results of deciding the number of latent classes**

The decision concerning the number of latent classes is based on a lower value of information criteria for AIC, BIC and adjusted BIC indices, or a nominal value of .05 for the VLMR, LMR and BLRT tests. First, the right two-class solution against the wrong one-class solution is compared in section 6.2.1. The results of model A.8 are first presented. Then, as in the case of results concerning the problems in estimation, the effect of reliability (A.8 vs. A.5), the effect of number of additional measurements (A.5 vs. A.5\*), and the effect of model constructs (C.8 vs. A.8, B8 vs. A.8 and B.5 vs. A.5) are examined in the context of deciding the number of latent classes.

Second, the results examining the wrong three-class solution against the right two-class solution in model A.8 are presented in section 6.2.2. These results are computed using Mplus version 4.2 with 10 or 500 random starting values.

### **6.2.1. Results of comparing two-class solution versus one-class solution**

This section presents the results concerning the proportion of decisions concluding to the right two-class solution instead of the wrong one-class solution. First, in section 6.2.1.1., the results using AIC, BIC and aBIC are presented. In section 6.2.1.2. the results using the VLMR and LMR tests are presented and also additional results using BLRT. These results are compared to the empirical distribution in section 6.2.1.3 and, finally, these results are summarized in section 6.2.1.4.

#### **6.2.1.1. Results of AIC, BIC and aBIC**

##### *The effect of sample size and SMD*

The next results describe the ability of AIC, BIC and adjusted BIC indices to detect the right number of latent classes in models A.8, A.5, A.5\*, B.8, B.5 and C.8. The number of latent classes is based on a lower value of information criteria

for AIC, BIC and the adjusted BIC indices. The decision is made between one or two classes. The focus will be in the cases, where the proportion of right decisions comes up to .70, which is the lowest acceptable line for successful use of any index or test. The estimates needed for sample sizes to achieve .80 and .90 lines are also informative, because these lines could be used as thresholds, when planning sample sizes in behavioral studies.

The results of proportions using AIC, BIC and aBIC are presented in Table 6.4. From this table one can see how the proportion of right decisions changes when the sample size or SMD increases. Based on the values in Table 6.4, some figures and linear approximations are used to evaluate the needed sample size to achieve .70, .80 or .90 proportions.

The results for model A.8, A.5, A.5\*, B.8, B.5 and C.8 show that, when SMD is 1, proportions for AIC and BIC are unchangeable and small with respect to  $n$ . Proportions for AIC varies between .186 and .278 and for BIC between .001 and .039. For aBIC, proportions decrease similarly in all models from .724 - .741 to .027 - .036. Next, the more informative results, when  $SMD \geq 2$ , are presented.

#### *Model A.8 - AIC*

When SMD is 2, the proportion of right decisions using AIC increases from .348 to .546 when the sample size increases from 50 to 200, and the proportion is approximately .70, .80 or .90 when the sample size is greater than 360, 450 or 670, respectively, as can be seen from Figure 6.1. When SMD is 3, the proportion for AIC is .703 when  $n = 50$ , and is approximately .80 or .90 when the sample size is greater than 75, or 95, respectively. When SMD is 4, the proportion for AIC is .977 when  $n = 50$  and 1.0 when  $n \geq 100$ . When SMD is 5, the proportion is 1.0 when  $n \geq 50$ .

#### *Model A.8 - BIC*

When SMD is 2, the proportion of the right decisions using BIC increases from .055 to .442 when the sample size increases from 50 to 1000. When SMD is 3, the proportion increases from .263 to .463 when the sample size increases from 50 to 100 and, as can be seen from Figure 6.2, the sample size needed to achieve .70, .80 or .90 proportion is about 170, 190 or 330, respectively. When SMD is 4, the proportion is .799 when  $n = 50$ , and the needed sample size to achieve a .90 proportion range is 80. When SMD is 5, the proportion is greater than .994 when  $n \geq 50$ .

Table 6.4. The proportions of the right two-latent-class decisions based on AIC, BIC and aBIC criteria in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8			A.5			A.5*			B.8			B.5			C.8		
n	SMD	AIC	BIC	aBIC	AIC	BIC	aBIC	AIC	BIC	aBIC	AIC	BIC	aBIC	AIC	BIC	aBIC	AIC	BIC	aBIC
50	1	.270	.036	.737	.278	.038	.736	.253	.032	.724	.270	.036	.737	.278	.038	.736	.275	.039	.741
100	1	.232	.011	.399	.238	.013	.403	.210	.010	.377	.232	.011	.399	.238	.013	.403	.228	.011	.400
200	1	.221	.005	.192	.218	.004	.188	.191	.003	.167	.221	.005	.192	.218	.004	.188	.213	.004	.185
500	1	.228	.001	.070	.210	.001	.064	.186	.001	.057	.228	.001	.070	.210	.001	.064	.210	.001	.066
1000	1	.241	.001	.036	.217	.001	.029	.197	.000	.027	.241	.001	.036	.217	.001	.029	.220	.001	.031
50	2	.348	.055	.807	.312	.047	.780	.284	.039	.757	.346	.058	.801	.314	.048	.776	.313	.047	.784
100	2	.404	.037	.593	.317	.023	.506	.309	.021	.484	.404	.037	.584	.331	.023	.514	.318	.020	.506
200	2	.546	.046	.508	.396	.018	.357	.411	.020	.372	.549	.045	.510	.398	.018	.360	.387	.017	.354
500	2	.847	.139	.644	.616	.031	.352	.685	.056	.432	.852	.143	.654	.628	.034	.364	.595	.030	.333
1000	2	.986	.442	.881	.859	.102	.518	.921	.199	.665	.987	.463	.879	.862	.111	.536	.845	.089	.492
50	3	.703	.263	.958	.543	.143	.906	.566	.164	.914	.663	.230	.949	.494	.115	.888	.534	.136	.901
100	3	.914	.463	.962	.745	.201	.866	.794	.274	.894	.879	.389	.947	.659	.142	.803	.731	.193	.858
200	3	.995	.819	.995	.936	.428	.925	.970	.580	.964	.990	.743	.988	.873	.275	.853	.929	.400	.917
500	3	1.0	.999	1.0	1.0	.930	.999	1.0	.983	1.0	1.0	.999	1.0	.998	.790	.986	1.0	.917	.996
1000	3	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.995	1.0	1.0	1.0	1.0
50	4	.977	.799	.999	.889	.536	.991	.925	.634	.995	.955	.720	.998	.782	.364	.973	.887	.507	.991
100	4	1.0	.982	1.0	.992	.841	.998	.997	.921	.999	.999	.957	1.0	.956	.613	.985	.989	.819	.996
200	4	1.0	1.0	1.0	1.0	.993	1.0	1.0	.998	1.0	1.0	1.0	1.0	.999	.930	.999	1.0	.991	1.0
500	4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1000	4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
50	5	1.0	.994	1.0	.996	.927	1.0	.998	.965	1.0	1.0	.980	1.0	.963	.753	.998	.993	.922	1.0
100	5	1.0	1.0	1.0	1.0	.999	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.999	.970	1.0	1.0	.997	1.0
200	5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
500	5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1000	5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

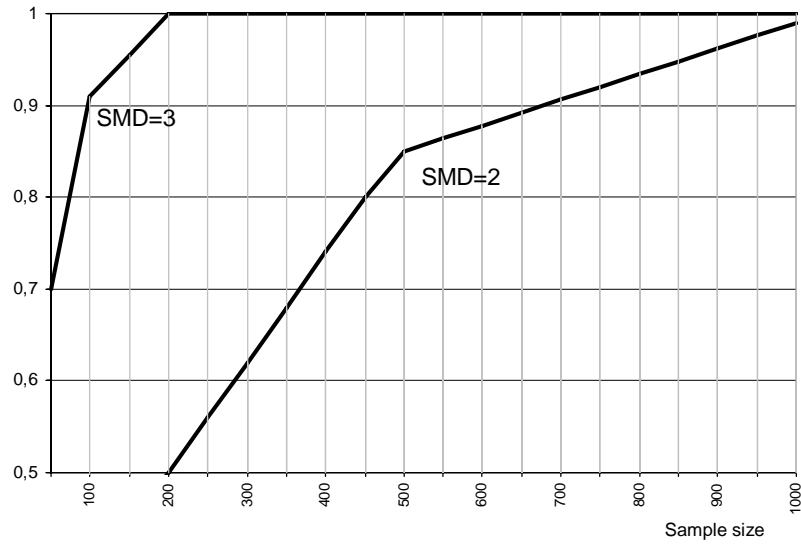


Figure 6.1. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using AIC criteria in model A.8 as a function of sample size,  $SMD = 2$  or  $3$ .

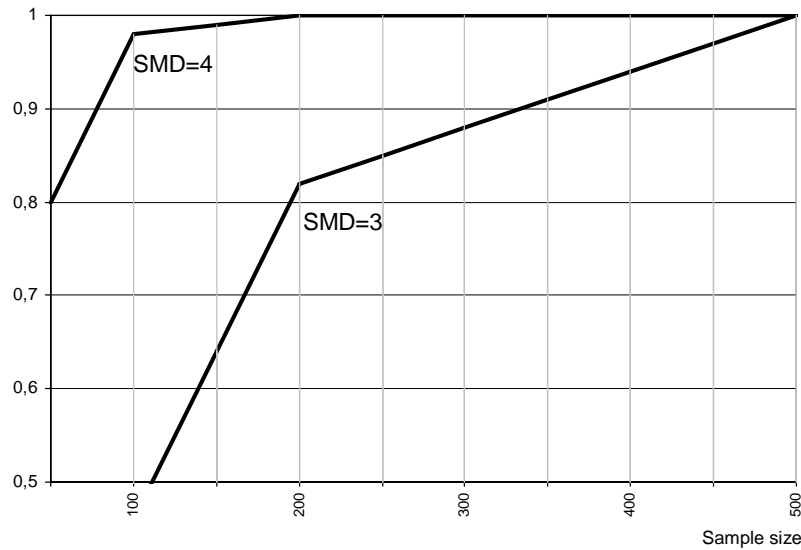


Figure 6.2. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using BIC criteria for model A.8 as a function of sample size,  $SMD = 3$  or  $4$ .

### Model A.8 - aBIC

When SMD is 1, the proportion of the right decisions using aBIC decreases from .737 to .036 when the sample size increases from 50 to 1000. This suspicious decrease is seen also when SMD is 2. When the sample size increases from  $n = 50$  to  $n = 200$  (see Figure 6.3), the proportion decreases from .807 to .508 and, after that, the sample size needed to achieve a .70, .80 or .90 proportion line is about  $n = 620, 840$  or  $> 1000$ , respectively. When  $SMD \geq 3$ , the proportion using aBIC is in most cases 1.0 and its value is at minimum .958.

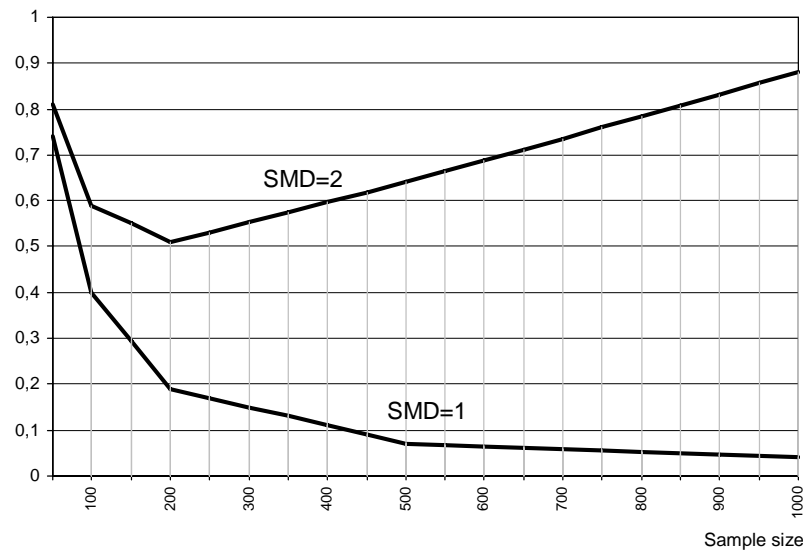


Figure 6.3. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using aBIC criteria for model A.8 as a function of sample size, SMD = 1 or 2.

### The effect of reliability on the deciding number of latent classes

#### Model A.5 vs. model A.8 - AIC

When using AIC, the effect of low reliability (A.5 vs. A.8) on the deciding number of latent classes is strong and is most obvious when SMD is 2 or 3. As can be seen from Table 6.4. and Figure 6.4., when SMD is 2, the proportion of the right two-class solutions in model A.5 using AIC is .70 or .80 when  $n \geq 675$  or  $880$ , respectively, which  $n$  is 315 or 430 greater, respectively, than in model A.8. When  $n = 1000$ , the proportion in model A.5 is .859. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion is about 90, 130 or 180, respectively, which  $n$  is more than 40, 55 or 85 greater, respectively, than in model A.8. When SMD is 4, the sample size needed to achieve a .70 or .80 proportion in model A.5

is lower than 50 as is in model A.8, whereas the sample size needed to achieve a .90 proportion is  $n \geq 55$ , which  $n$  is at least 5 greater than in model A.8. As in model A.8, when SMD is 5, in model A.5., the sample size needed to achieve a .70, .80 or .90 proportion is lower than 50.

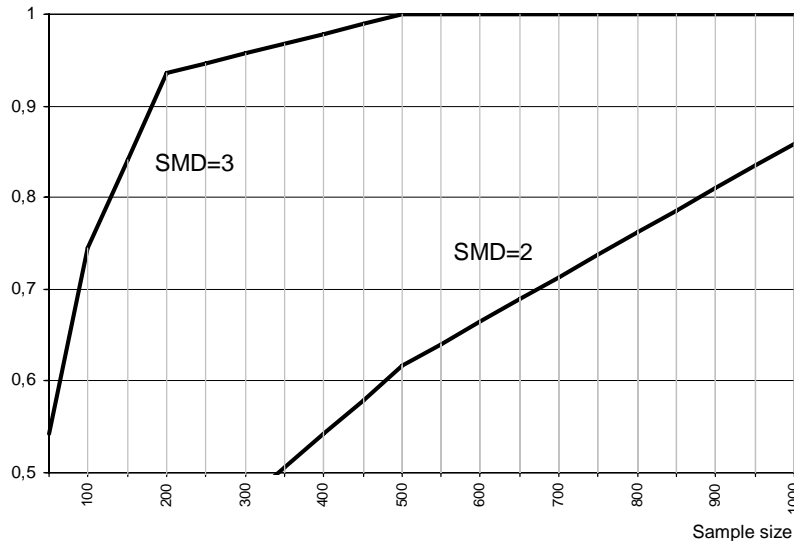


Figure 6.4. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using AIC criteria for model A.5 as a function of sample size, SMD = 2 or 3.

*Model A.5 vs. model A.8 – BIC*

When SMD is 2, the proportion of the right decisions using BIC is even lower in model A.5, i.e. between .047 and .102, than in model A.8, as can be seen from Table 6.4. When SMD is 3, the sample size needed to achieve .70, .80 or .90 proportions in model A.5 (see Figure 6.5) are about 365, 420 or 480, respectively, which  $n$  is 200, 225 or 145 greater than in model A.8. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 75, 95 or 140, respectively, which  $n$  is more than 25, 45 or 60 greater than in model A.8. When SMD is 5, the sample size needed to achieve a .70, .80 or .90 proportion is in model A.5 lower than 50 similar to model A.8.

*Model A.5 vs. model A.8 – aBIC*

When SMD=1, as is model A.8, the proportion to conclude to the two-class solution using aBIC decreases when the sample size increases in model A.5 (see Table 6.4). The effect of reliability is seen when SMD is 2, in which case the suspicious decrease of proportion is even stronger in model A.5 than in model A.8.

When  $SMD \geq 3$ , the proportion in model A.5 is over .90, except when SMD is 3 and  $n = 100$ , in which case the proportion is .87.

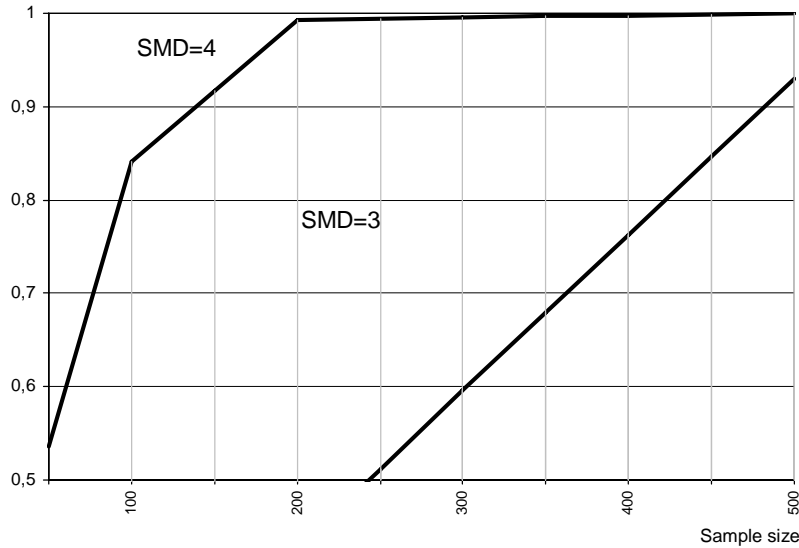


Figure 6.5. The proportion of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using BIC criteria for model A.5 as a function of sample size,  $SMD = 3$  or  $4$ .

*The effect of additional measurement points on the deciding number of latent classes*

Next, the effect of additional measurement points is examined comparing model A.5\* with model A.5. The results are also compared with those in model A.8. Model A.5\* has three additional measurement points compared with model A.5. Although SMD of latent factors is the same for both models (A.5 and A.5\*),  $SMD(y)$  in model A.5\* is greater than in model A.5, and lower than in model A.8.

*Model A.5\* vs. model A.5 – AIC*

As can be seen from Table 6.4, when using AIC, proportions of the right decisions in model A.5\* are close to the respective proportions in model A.5 when SMD is 1. In model A.5\*, when SMD is 2, the proportion (see Figure 6.6) increases from .284 to .685, when the sample size increases from 50 to 500 and the sample size needed to achieve a .70 or .80 proportion line is 530 or 745, respectively, which  $n$  is 145 or 135 less, respectively, than in model A.5 and 315 or 430 more than in model A.8. The sample size needed to achieve the .90 proportion line is about  $n =$



950, which  $n$  is 250 more than in model A.8. When SMD is 3, the proportion increases from .566 to .794 when the sample size increases from 50 to 100 and the sample size needed to achieve a .70, .80 or .90 proportion line is 80, 105 or 160, respectively, which  $n$  is 10, 25 or 20 less, respectively, than in the model A.5 and at least 40, 55 or 85 more than in model A.8. When SMD is 4 or 5, the proportion in model A.5\* is as high as it is in models A.5 and A.8, and is over .90 when  $n \geq 50$ .

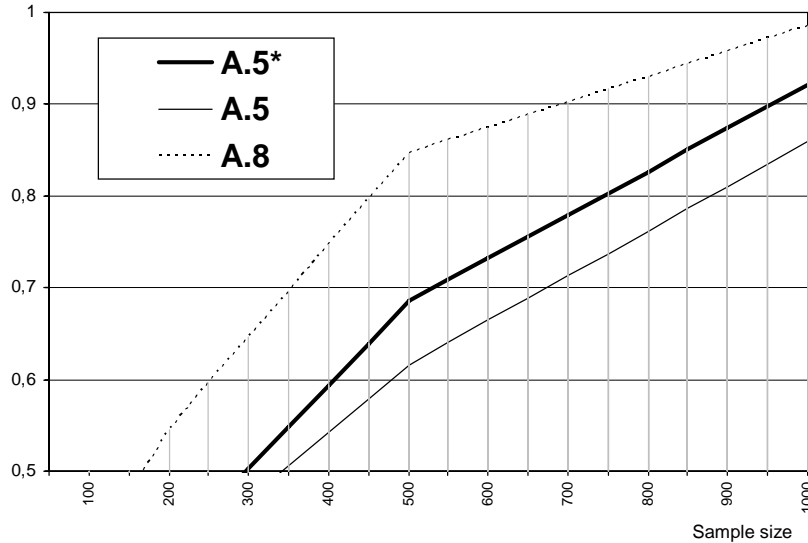


Figure 6.6. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using AIC index for models A.5\*, A5 and A.8 as a function of sample size, SMD = 2.

*Model A.5\* vs. model A.5 – BIC*

When using BIC, the proportion is low in model A.5\* as in model A.5 when SMD is 1 or 2. As can be seen from Figure 6.7, when SMD is 3, the sample size needed in model A.5\* to achieve a .70, .80 or .90 proportion line is about 290, 365 or 440, respectively, which  $n$  is 75, 55 or 40 less than in the case of model A.5, and 125, 170 or 105 more than in the model A.8. When SMD is 4, the proportion using BIC is .634 when  $n = 50$ , and the sample size needed to achieve a .70, .80 or .90 proportion line is 60, 80, or 95, respectively, when  $n$  is 15, 15 or 45 less, respectively, than in model A.5 and at least 10, 30 or 15 more, respectively, than in model A.8. When SMD is 5, the proportion is over .90 when  $n \geq 50$  as in model A.5.

*Model A.5\* vs. model A.5 - aBIC*

As can be seen from Table 6.4., the proportion of the right two-class solutions in model A.5\* using aBIC is similar to model A.5 when SMD is 1 or 2. When

SMD  $\geq 3$ , the proportion is over .90, except when SMD is 3 and  $n = 100$  in which case the proportion is .894.

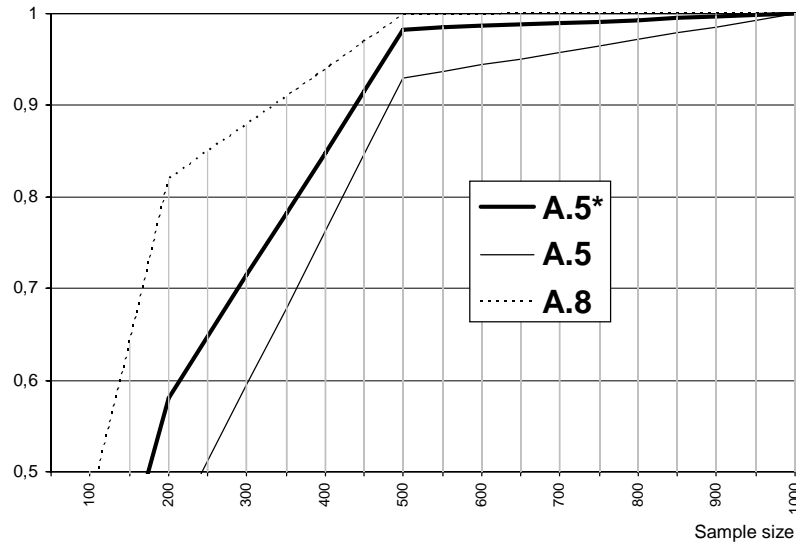


Figure 6.7. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using BIC index for models A.5\*, A.5 and A.8 as a function of sample size, SMD = 3.

*The effect of construct on the deciding number of latent classes*

The effect of construct on the decision of the number of latent classes when using AIC, BIC or aBIC, is, in average, weak when comparing model B.8 with A.8 and model B.5 with A.5 (see Table 6.4). The largest effects are seen when SMD is 3.

*Model C.8 vs. model A.8 – AIC*

When using AIC, the proportions to conclude to the true two-classes versus the wrong one-class in model C.8 are comparable to those in model A.8 in the sense of SMD. Despite this, the proportions in model C.8 are clearly lower than in model A.8, but are very similar to the proportions in model A.5. When SMD is 2, the sample size needed to achieve a .70 or .80 proportion in model C.8 is 710 or 910, respectively, which  $n$  is 350 or 460 more than in model A.8 and 35 or 30 more than in model A.5. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model C.8 is 90, 135 or 185, respectively, which  $n$  is at least 40, 60 or 90 more than in model A.8 and 0 - 5 more than in model A.5. When SMD is 4 or 5, the proportion in model C.8 is greater than .90 when  $n \geq 50$ , as it is in model A.8, except when SMD is 4, the proportion in model C.8 is .90 when  $n \geq 55$  similar to model A.5.

The results regarding the proportions in model C.8 when using AIC, are more similar to the respective proportions in model A.5 than to model A.8. This can be explained by the fact that the  $SMD(\mathbf{y})$  in model A.5 is equal with  $SMD(\mathbf{y})$  in model C. These results suggest that the proportion to detect two classes versus one class is related more to the  $SMD(\mathbf{y})$  than to the SMD of latent components.

#### *Model C.8 vs. model A.8 – BIC*

When SMD is 2, the proportion of the right decisions in model C.8 using BIC is low .047 -.089, as in model A.8 or model A.5. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model C.8 is 375, 430 or 490, respectively, which  $n$  is 210, 235 or 155, respectively, more than in model A.8 and 10 more than in model A.5. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion in model C.8 is 80, 95 or 145, which  $n$  is at least 30, 45 or 65, respectively, more than in model A.8 and 0-5 more than in model A.5. As in model A.8 or A.5, when SMD is 5, the proportion is over .90 when  $n \geq 50$ .

#### *Model C.8 vs. model A.8 – aBIC*

When SMD is 2, the proportion in model C.8 using aBIC decreases from .784 to .492 when the sample size increases from 50 to 1000. When  $SMD \geq 3$ , as in A.5, the proportions are greater than .90, except when SMD is 3 and  $n = 100$ , in which case the proportion is .858.

#### *Model B.8 vs. model A.8 – AIC*

When using AIC, the proportions in model B.8 are very similar to that in model A.8. When SMD is 2, the sample size needed to achieve a .70, .80 or .90 proportion in model B.8 is 350, 450 or 680, respectively, which  $n$  is 10 less, similar, or 10 more, respectively, than in model A.8. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model B.8 is 60, 80 or 120, which  $n$  is at least 10, 15 or 25 more, respectively, than in model A.8. As in model A.8, in model B.8, the proportions using AIC are over .90 when  $SMD \geq 4$  and  $n \geq 50$ .

#### *Model B.8 vs. model A.8 – BIC*

When using BIC, the proportions in model B.8 are very similar to those in model A.8. When SMD is 2, the proportion using BIC is small and increases from .058 to .463 when the sample size increases from 50 to 1000. When SMD is 3, the sample size needed to achieve .70, .80 or .90 proportion in model B.8 is 190, 265 or 385 which  $n$  is 25, 70 or 50 more, respectively, than in model A.8. When SMD is 4, as in model A.8, the proportion using BIC is over .70 in model B.8 when  $n \geq 50$ . The sample size needed to achieve a .80 or .90 proportion in model B.8 is 65 or 90,

which  $n$  is 15 or 10 more, respectively, than in model A.8. As in model A.8, in model B.8, the proportions using BIC are over .90 when  $SMD = 5$  and  $n \geq 50$ .

*Model B.8 vs. model A.8 – aBIC*

When using aBIC, the proportions in model B.8 are very similar to those in model A.8. When SMD is 2, the proportion in model B.8 first decreases from .801 to .510 when the sample size increases from 50 to 200 and, after that, increases achieving a .70, .80, or .90 proportion when the sample size increases to 595, 805 or greater than 1000, respectively. As in model A.8, the proportion in model B.8 is over .90 when  $SMD \geq 3$  and  $n \geq 50$ .

*Model B.5 vs. model A.5 – AIC*

When using AIC, the proportions in model B.5 are very similar to those in model A.5. When SMD is 2, the sample size needed to achieve .70 or .80 proportion in model B.5 is 655 or 870, which  $n$  is 20 or 10 less, respectively, than in model A.5. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model B.5 is 120, 165 or 265, which  $n$  is 30, 35 or 85 more, respectively, than in model A.5. When SMD is 3, the proportion in model B.5 is .782 when  $n = 50$ , which  $n$  is .107 lower than in model A.5. The sample size needed to achieve a .80 or .90 proportion in model B.5 is 55 or 85, which  $n$  is at least 5 or 30 more, respectively, than in model A.5. As in model A.5, in model B.5, the proportion is greater than .90 when SMD is 5 and  $n \geq 50$ .

*Model B.5 vs. model A.5 – BIC*

When using BIC, the proportions in model B.5 are very similar to those in model A.5. When SMD is 2, the proportion in model B.5 is low varying between .048 - .111. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model B.5 is 450, 525 or 770, which  $n$  is 85, 105 or 290 more, respectively, than in model A.5. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion in model B.5 is 120, 160 or 190, which  $n$  is 45, 65 or 50 more, respectively, than in model A.5. When SMD is 5, the proportion in model B.5 is .753 when  $n = 50$ , which is .174 lower than in model A.5. The sample size needed to achieve a .80 or .90 proportion in model B.5 is 60 or 85, which  $n$  is at least 10 or 35 more, respectively, than in model A.5. As in model A.5, in model B.5, the proportion is greater than .90 when SMD is 5 and  $n \geq 50$ .

*Model B.5 vs. model A.5 – aBIC*

When using aBIC, the proportions in model B.5 are very similar to those in model A.5. When SMD is 2, the proportion in model B.5 first decreases from .776 to .360 when the sample size increases from 50 to 200 and, after that, increases to .536 when the sample size increases to 1000. When SMD is 3, the proportions are lower

in model B.5 than in model A.5, the proportion first decreasing from .888 to .803 when the sample size increases from 50 to 100 and, after that, increasing to .853 when the sample size increases to 200. The sample size needed to achieve a .90 proportion in model B.5 is 305, which  $n$  is 145 more than in model A.5. As in model A.5, in model B.5, the proportion is greater than .90 when SMD is 4 or 5 and  $n \geq 50$ .

### **6.2.1.2. Results of VLMR, LMR and BLRT**

The next results describe the ability of the VLMR and LMR tests to detect the right number of two-latent-classes versus the wrong one-class in nominal .05 level in models A.8, A.5, A.5\*, B.8, B.5 and C.8. As for AIC, BIC and aBIC, the focus will be in the cases where the proportion of right decisions comes up to .70, .80 and .90. Addition to this, some results using BLRT are also presented in the text.

The results presented in Table 6.5 for models A.8, A.5, A.5\*, B.8, B.5 and C.8 show that, when SMD is 1, the proportion of the right number of latent classes using the VLMR or LMR test is unchangeable and small with respect to  $n$ . The proportion varies between .027 and .090 when using VLMR and between .021 - .086 when using LMR. When  $SMD \geq 2$ , the proportions increase when the sample size or SMD increases. In the following sections, these results are presented in more detail.

#### *Model A.8 - VLMR*

When SMD is 2, the proportion using VLMR increases from .039 to .481 when the sample size increases from 50 to 500 and the proportion is .826 when  $n = 1000$  (see Figure 6.8). The sample size needed to achieve a .70 proportion is then 820. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion is 160, 185 or 300, respectively. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 70, 80 or 95, respectively. When SMD is 5, the proportion is over .90 when  $n \geq 50$ .

#### *Model A.8 - LMR*

When SMD is 2, the proportion using LMR increases from .030 to .454 when the sample size increases from 50 to 500 and the proportion is .810 when  $n = 1000$  (see Figure 6.9). The sample size needed to achieve a .70 or .90 proportion is 845 or 985, respectively. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion is 170, 190 or 320, respectively. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 70, 85 or 95, respectively. When

SMD is 5, the proportion is .882 when n = 50 and the sample size needed to achieve the .90 proportion is 60.

Table 6.5. The proportions of the right two-latent-class versus wrong one-class decisions based on the VLMR and LMR tests in models A.8, A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8		A.5		A.5*		B.8		B.5		C.8	
n	SMD	VLMR	LMR	VLMR	LMR	VLMR	LMR	VLMR	LMR	VLMR	LMR	VLMR	LMR
50	1	.032	.024	.031	.024	.027	.021	.032	.024	.031	.024	.029	.022
100	1	.043	.033	.040	.033	.045	.037	.043	.033	.040	.033	.044	.035
200	1	.054	.044	.042	.045	.062	.053	.054	.044	.042	.045	.054	.047
500	1	.075	.065	.069	.061	.075	.064	.075	.065	.069	.061	.077	.066
1000	1	.090	.081	.085	.074	.086	.076	.090	.081	.085	.074	.085	.076
50	2	.039	.030	.034	.025	.028	.022	.039	.030	.031	.024	.033	.027
100	2	.075	.059	.053	.042	.057	.045	.077	.062	.054	.043	.049	.038
200	2	.175	.148	.099	.082	.119	.102	.173	.148	.102	.086	.099	.080
500	2	.481	.454	.254	.231	.384	.351	.502	.472	.272	.248	.245	.220
1000	2	.826	.810	.525	.499	.753	.729	.836	.822	.545	.521	.504	.476
50	3	.150	.121	.079	.061	.083	.061	.132	.106	.060	.045	.077	.059
100	3	.463	.421	.238	.203	.327	.282	.405	.366	.178	.150	.221	.190
200	3	.850	.834	.585	.550	.790	.762	.818	.792	.465	.427	.570	.536
500	3	.999	.998	.969	.964	.999	.999	.997	.996	.919	.910	.966	.962
1000	3	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.998	.997	1.0	1.0
50	4	.570	.520	.318	.270	.396	.340	.470	.417	.192	.159	.298	.252
100	4	.935	.923	.769	.735	.908	.889	.896	.876	.555	.517	.745	.713
200	4	.998	.998	.984	.981	.999	.999	.997	.997	.928	.919	.981	.979
500	4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1000	4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
50	5	.904	.882	.721	.675	.840	.800	.835	.802	.486	.431	.699	.652
100	5	.997	.997	.978	.973	.999	.998	.992	.991	.901	.884	.976	.972
200	5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.998	.997	1.0	1.0
500	5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1000	5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Note. Bootstrapped loglikelihood ratio tests were done only for some of the above models because of heavy calculation (e.g., when SMD = 2 and n = 500 calculation time was 33 h 7 min).

*Model A.8 - BLRT*

The bootstrapped likelihood ratio is calculated in only three situations. The reason for that is that it requires very heavy computation. In model A.8, when SMD is 2

and  $n = 500$ , the proportion using BLRT is .535. When SMD is 3 and  $n = 200$ , the proportion is .601, and when SMD is 4, and  $n = 50$ , the proportion is .768. These proportions are clearly higher than those found when using the VLMR or LMR tests.

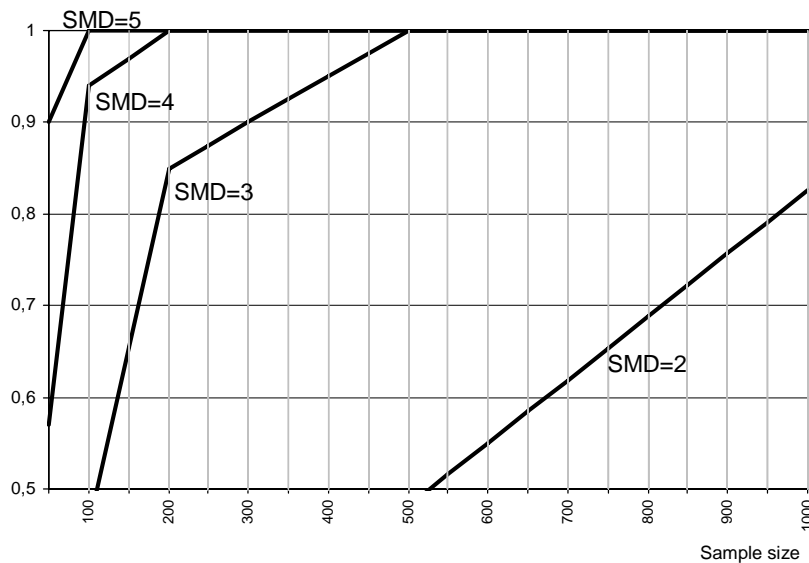


Figure 6.8. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using the VLMR test for model A.8 as a function of sample size and SMD.

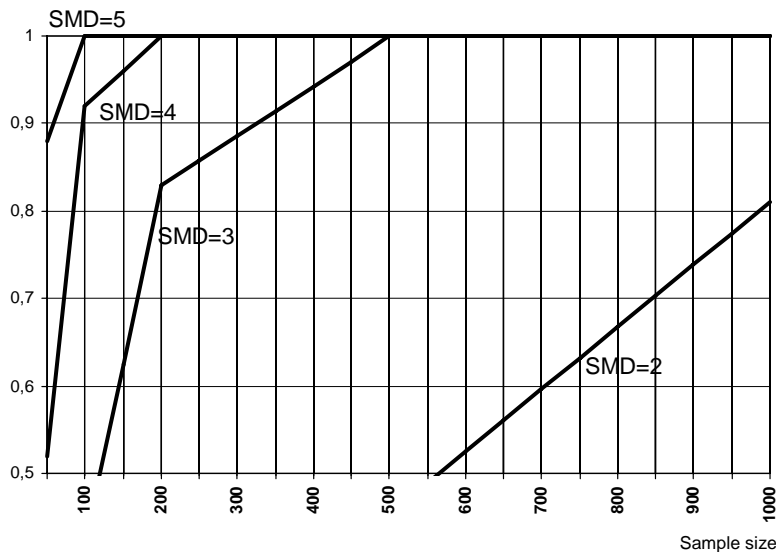


Figure 6.9. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using the LMR test for model A.8 as a function of sample size and SMD.

*The effect of reliability on the deciding the number of latent classes*

The effect of reliability on the proportion of two-class solutions using the VLMR and LMR tests (see Table 6.5) is similar to using BIC. The proportions using the VLMR or LMR tests are low in model A.5 when SMD is 1 or 2. In model A.8, when SMD is 2 and the sample size is 1000, the proportion is .83 or .81 using the VLMR or LMR tests, respectively, whereas in model A.5, the proportion is .53 using the VLMR test and .50 using the LMR test. The effect of reliability (A.5 vs. A.8) using VLMR or LMR is most obvious when SMD is 3 or 4. This is presented in the following text.

*Model A.5 vs. model A.8 - VLMR*

As can be seen from Figure 6.10, when SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion using VLMR in model A.5 is 290, 370 or 445, respectively, which n is 130, 185 or 145 more than in model A.8. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 90, 115 or 160, respectively, which n is 20, 35 or 65 more than in model A.8. When SMD is 5, the proportion is .721 when n = 50 and the sample size needed to achieve a .80 or .90 proportion is 65 or 85, respectively, which n is at least 15 or 35 more, respectively, than in model A.8.

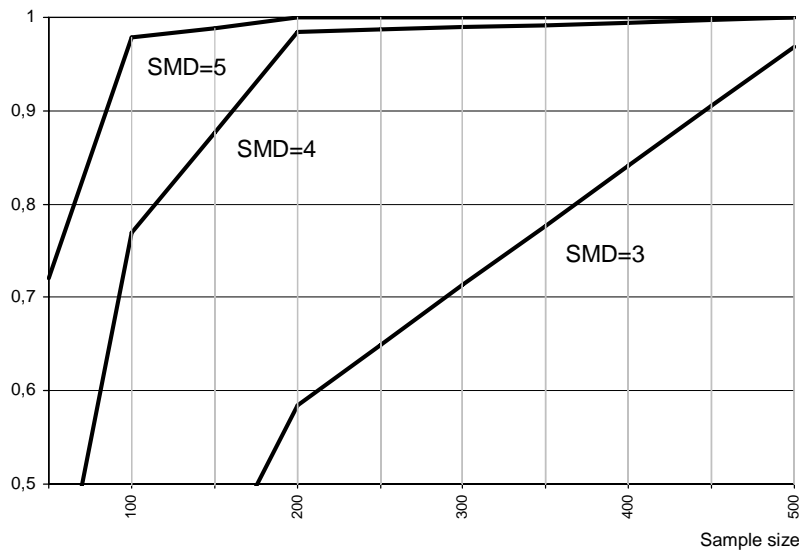


Figure 6.10. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using the VLMR test for model A.5 as a function of sample size, SMD = 3, 4 or 5.



#### *Model A.5 vs. model A.8 - LMR*

When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion using LMR in model A.5 is 310, 380 or 455, respectively, which n is 140, 190, 135 more than in model A.8. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion in model A.5 is 95, 125 or 165, respectively, which n is 25, 40 or 70 more than in model A.8. When SMD is 5, the sample size needed to achieve a .70, .80 or .90 proportion is 55, 70 or 90, respectively, which n is at least 5, 20 or 30 more than in model A.8.

#### *The effect of additional measurement points on deciding the number of latent classes*

As can be seen from Table 6.5, the proportions using the VLMR or LMR tests in model A.5\* are greater than the proportions in model A.5.

#### *Model A.5\* vs. model A.5 – VLMR*

When SMD is 2, the proportion in model A.5\* using VLMR increases from .028 to .753 when the sample size increases from 50 to 1000 and the proportion is .70 when  $n \geq 930$ . As can be seen from Figure 6.11., the effect of additional measurement points on the proportion is clear and the proportion line in model A.5\* is near to the proportion line of model A.8. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model A.5\* is 180, 215 or 360, respectively, which n is 110, 155 or 85 less, respectively, than in model A.5 and 20, 30 or 60 more, respectively, than in model A.8. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion in model A.5\* is 80, 90 or 100, respectively, which n is 10, 25 or 60 less, respectively, than in model A.5 and 10, 10 or 5 more, respectively than in model A.8. When SMD is 5, the proportion is .840 when  $n = 50$ , which n is .119 greater than in model A.5 and .064 lower than in model A.8. The sample size needed to achieve a .90 proportion in model A.5\* is 70, which n is 15 less than in model A.5, and at least 20 more than in model A.8.

#### *Model A.5\* vs. model A.5 – LMR*

When SMD is 2, the proportion in model A.5\* using LMR increases from .022 to .729 when the sample size increases from 50 to 1000 and is .70 when  $n \geq 960$ . When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model A.5\* is 185, 250 or 375, respectively, which n is 125, 130 or 80 less, respectively, than in model A.5 and 15, 60 or 55 more, respectively than in model A.8. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 85, 90 or 110, respectively, which n is 10, 35 or 45 less, respectively, than in model A.5 and 15, 5 or 15 more, respectively than in model A.8. When SMD is 5, the proportion in model A.5\* is .800 when  $n = 50$ , which proportion is

.125 greater than in model A.5 and .082 lower than in model A.8. The sample size needed to achieve a .90 proportion in model A.5\* is 75, which n is 15 less than in model A.5 and 15 more than in model A.8.

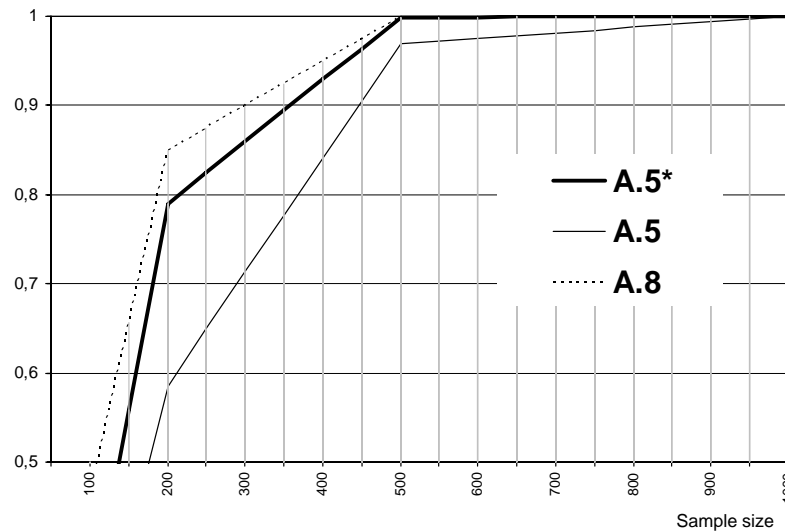


Figure 6.11. The proportions of 10000 replications concluding to the right two-latent-class solution instead of to the wrong one-class solution using the VLMR test for models A.5\*, A.5 and A.8 as a function of sample size, SMD=3.

*The effect of construct on deciding the number of latent classes*

When using VLMR or LMR, the effect of construct on the decision of number of latent classes is, on average, weak when comparing model B.8 with A.8, and model B.5 with A.5 (see Table 6.5). The largest differences between pairs of models are seen when SMD is 3.

*Model C.8 vs. model A.8 - VLMR*

When using VLMR, the proportions in model C.8 are clearly lower than in model A.8. When SMD is 2, the proportion in model C.8 increases from .033 to .504 when the sample size increases from 50 to 1000 and, when n = 1000, the proportion is .322 lower in model C.8 than in model A.8. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model C.8 is 300, 375 or 450, respectively, which n is 140, 190 or 150 more, respectively, than in model A.8. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 95, 125 or 165, respectively, which n is 25, 45 or 70 more, respectively, than in model A.8. When SMD is 5, the sample size needed to

achieve a .70, .80 or .90 proportion is 50, 70 or 85, respectively, which n is at least 0, 20 or 35 more, respectively, than in model A.8.

*Model C.8 vs. model A.8 - LMR*

When using LMR, the proportions in model C.8 are clearly lower than in model A.8. When SMD is 2, the proportion in model C.8 increases from .027 to .476 when the sample size increases from 50 to 1000 and, when  $n = 1000$ , the proportion is .334 lower in model C.8 than in model A.8. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model C.8, is 315, 385 or 455, respectively, which n is 145, 195 or 135 more, respectively, than in model A.8. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion, is 100, 135 or 170, respectively, which n is 30, 50 or 75 more, respectively, than in model A.8. When SMD is 5, the sample size needed to achieve a .70, .80 or .90 proportion, is 60, 75 or 90, respectively, which n is at least 10, 15 or 30 more, respectively, than in model A.8.

*Model B.8 vs. model A.8 - VLMR*

When SMD is 2, the sample size needed to achieve a .70 or .80 proportion in model B.8 using VLMR, is 800 or 945, respectively, which n is 20 or at least 55 less, respectively, than in model A.8. When  $n = 1000$ , the proportion in model B.8 is .836, which is .010 greater than in model A.8. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model B.8 is 170, 195 or 335, respectively, which n is 10, 10 or 35 more, respectively, than in model A.8. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 75, 90 or 105, respectively, which n is 5, 10 or 10 more, respectively, than in model A.8. When SMD is 5, the proportion in model B.8 is .835 which is .069 lower than in model A.8. The sample size needed to achieve a .90 proportion in model B.8 is 70, which is at least 20 greater than in model A.8.

*Model B.8 vs. model A.8 - LMR*

When SMD is 2, the sample size needed to achieve a .70 or .80 proportion in model B.8 using LMR, is 825 or 970, respectively, which n is 20 or 15 less, respectively, than in model A.8. When  $n = 1000$ , the proportion in model B.8 is .822, which is .012 greater than in model A.8. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model B.8 is 180, 210 or 360, respectively, which n is 10, 20 or 40 more, respectively, than in model A.8. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 80, 90 or 120, respectively, which n is 10, 5 or 25 more, respectively, than in model A.8. When SMD is 5, the proportion in model B.8, is .802 which is .080 lower than in model A.8. The sample size needed to achieve a .90 proportion in model B.8, is 75 which is 15 more than in model A.8.

### *Model B.5 vs. model A.5 - VLMR*

The proportions in model B.5 using VLMR are clearly lower than in model A.5 when  $SMD \geq 3$ . When SMD is 2, the proportion in model B.5 increases from .031 to .545 when the sample size increases from 50 to 1000 and, when  $n = 1000$ , the proportion in model B.5 is .020 greater than in model A.5. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model B.5 is 355, 420 or 485, respectively, which  $n$  is 65, 50 or 40 more, respectively, than in model A.5. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 140, 165 or 190, respectively, which  $n$  is 50, 50 or 30 more, respectively, than in model A.5. When SMD is 5, the sample size needed to achieve a .70, .80 or .90 proportion in model B.5 is 75, 90 or 100, respectively, which  $n$  is at least 25, 25 or 15 more, respectively, than in model A.5.

### *Model B.5 vs. model A.5 - LMR*

The proportions in model B.5 using LMR are clearly lower than in model A.5 when  $SMD \geq 3$ . When SMD is 2, the proportion in model B.5 increases from .024 to .521 when the sample size increases from 50 to 1000 and, when  $n = 1000$ , the proportion in model B.5 is .022 greater than in model A.5. When SMD is 3, the sample size needed to achieve a .70, .80 or .90 proportion in model B.5 is 370, 430 or 495, respectively, which  $n$  is 60, 50 or 40 more, respectively, than in model A.5. When SMD is 4, the sample size needed to achieve a .70, .80 or .90 proportion is 145, 170 or 195, respectively, which  $n$  is 50, 45 or 30 more, respectively, than in model A.5. When SMD is 5, the sample size needed to achieve .70, .80 or .90 proportion in model B.5 is 80, 90 or 115, respectively, which  $n$  is 25, 20 or 25 more, respectively, than in model A.5.

### **6.2.1.3. Evaluating information criteria to produce wrong two-class solution versus right one-class solution**

To shed light on the results concerning information criteria presented above, penalties of different information criteria are next compared. This examination is done because the only difference between criteria is that they have different penalty functions. Deciding the number of classes is based on the smaller value of information criteria between two competitive, i.e.  $k$  and  $k+1$  class, models. This comparison can be expressed as a difference between information criteria in the two models.

$$AIC_{k+1} - AIC_k = [-2(\log L_{k+1} - \log L_k)] + [2(\mathbf{P}_{k+1} - \mathbf{P}_k)],$$

$$BIC_{k+1} - BIC_k = [-2(\log L_{k+1} - \log L_k)] + [\log(n)(\mathbf{P}_{k+1} - \mathbf{P}_k)],$$

$$aBIC_{k+1} - aBIC_k = [-2(\log L_{k+1} - \log L_k)] + \left[ \log\left(\frac{n+2}{24}\right)(P_{k+1} - P_k) \right],$$

where

$P_{k+1}$  and  $P_k$  are freely estimated parameters in  $k+1$  and  $k$  class models, respectively.

As can be seen from the equations presented above, the common  $-2$  times likelihood ratio term  $[-2(\log L_{k+1} - \log L_k)]$  is compared to the respective penalty (marked with bold), which is specific to each of the information criteria. The values of these penalties of information criteria are presented in Table 6.6 below. If the abscissa of the common term is greater than the value of the specific term, the conclusion is that the data has  $k+1$  latent classes; otherwise the data has  $k$  classes.

To compare the sufficiency of penalties, the values respecting a certain cumulative percentage for empirical distribution of  $2(\log L_2 - \log L_1)$  are estimated using one class data. The data is generated by using model A.8 with class 1 parameter values and analyzed by fitting a two-class model using 500 starting values to find the maximum likelihood. This process is replicated 10000 times. The model estimation in each of the replications is completed without any convergence problem. The results of these simulations for  $2(\log L_2 - \log L_1)$  representing a certain percentage of cumulative distribution are presented in Table 6.6.

Table 6.6. The penalties of AIC, BIC and aBIC and values of  $2(\log L_2 - \log L_1)$  representing 70, 80, 90, 95, 99 or 99.9 percentages of cumulative distribution when using data based on model A.8 with class\_1 parameter values.

n	Penalties			Cumulative percentages and respective values for $2(\log L_2 - \log L_1)$					
	AIC	BIC	aBIC	<b>70 %</b>	<b>80 %</b>	<b>90 %</b>	<b>95 %</b>	<b>99 %</b>	<b>99.9 %</b>
50	6	11.74	2.32	7.30	8.47	10.40	12.35	15.98	21.98
100	6	13.82	4.34	6.63	7.79	9.60	11.38	15.27	20.30
200	6	15.90	6.39	6.28	7.44	9.22	10.79	14.57	20.91
500	6	18.64	9.12	6.10	7.17	8.85	10.59	14.25	18.34
1000	6	20.72	11.20	6.04	7.06	8.79	10.45	14.14	19.47

As can be seen from Table 6.6, 95 percent of  $2(\log L_2 - \log L_1)$  values are lower than 12.35, 11.38, 10.79, 10.59 or 10.45 when  $n = 50, 100, 200, 500$  or 1000, respectively. These critical values are comparable to the .05 type I error rate and are 1.74 - 2.06 times larger than penalty values of AIC, and 0.50 – 1.05 times the penalty values of BIC. The penalty value of aBIC is small when  $n = 50, 100$  or 200 but is slightly lower when  $n = 500$ , and slightly greater when  $n = 1000$ , than the 95 % cumulative value of  $2(\log L_2 - \log L_1)$ . The results indicate that the penalty value of AIC is small, too often producing the wrong two-class solutions, whereas the penalty value of BIC is high producing lower than 0.1 percent wrong two-class solutions, when  $n = 500$  or 1000. The penalty of aBIC is very small with small sample sizes, but increases and is useful when the sample size is 500-1000.

The results for AIC, BIC and aBIC, as well as for the VLMR and LMR tests with .05 nominal level, using above data, in which the proportions of the wrong two-class solutions instead of the right one-class solutions are calculated for over 10000 replications, are presented in Table 6.7. Also, the proportions using the ordinary likelihood ratio test (OLRT) are presented.

Table 6.7. The proportions of 10000 replications concluding to the wrong two-class solution instead of to the right one-class solution in model A.8 using class 1 parameter values.

<b>n</b>	<b>AIC</b>	<b>BIC</b>	<b>aBIC</b>	<b>VLMR</b>	<b>LMR</b>	<b>OLRT</b>
50	.451	.063	.938	.071	.057	.253
100	.370	.020	.604	.102	.085	.198
200	.328	.007	.288	.135	.116	.173
500	.310	.001	.089	.167	.147	.152
1000	.303	.0004	.037	.190	.170	.148

Using AIC, the proportions are very high and the wrong two-class solution is produced in 30.3 - 45.1 percentages of cases. Using BIC, the wrong two-class solution instead of the right one-class solution is produced in 6.3 percentage of replications when  $n = 50$  and this percentage decreases to 2.0 or 0.7 when the sample size increases to 100 or 200, respectively. When the sample size is 500 or 1000, the proportions are very small, 0.1 or 0.04 percentages, respectively.

The proportions using aBIC are reverse for BIC: the proportion is high when the sample size is small and it decreases from .938 to .288 when the sample size

increases from 50 to 200. When the sample size is 500 or 1000, the proportions using aBIC are .089 or .037, respectively.

As can be seen from the results, the VLMR and LMR tests produce too often the wrong two-class solutions. The proportion is high .15 - .19 when  $n = 500-1000$ . The results mean that Type I error rates for these tests are high, even higher than when using the OLRT.

#### **6.2.1.4. Summary of the results comparing two-class solution versus one-class solution**

When combining the results of 6.2.1.1 and 6.2.1.2 concerning the right two-class solutions and respecting the evaluation of information criteria on section 6.2.1.3, the following conclusions can be made.

In all, the smallest sample sizes needed to achieve the .70, .80 or .90 proportion line when concluding to the right two-class solution instead of to the one-class solution is found for AIC and aBIC. These results are clearly seen from Tables 6.4 and 6.5 and Figures 6.12 and 6.13. Unfortunately, AIC with all sample sizes and aBIC with  $n \leq 200$  also produce a very high proportion concluding to the wrong two-class solution instead of to the right one-class solution (see Table 6.7). These results prevent the use of AIC in all, and also the use of aBIC when the sample size is small,  $n \leq 200$ .

In all, the highest samples sizes needed to achieve the .70, .80 or .90 proportion line when concluding to the right two-class solution instead of the one-class solution is found for BIC when  $SMD \leq 3$  and for VLMR when SMD is 4 or 5. These results are clearly seen from Tables 6.4 and 6.5 and Figures 6.12 and 6.13. When taking the results concerning the proportions concluding to the wrong two-class solution instead of the right one-class solution into account, it is shown that, when  $n = 50$ , BIC, VLMR and LMR are useful criteria to decide the number of latent classes. With this sample size, the .70 proportion line is achieved when SMD is 4 or 5, in which cases BIC needed a smaller sample size than VLMR or LMR. When the sample size is 100 or greater, the proportions of the wrong two-class solutions increase when using the VLMR and LMR tests, preventing their reliable use. In turn, when using BIC, the proportion decreases and is very small, when  $n$  is greater than 200. The results suggest that BIC is more useful than the VLMR or LMR tests when making decisions between the right two-class solutions and the wrong one-class solution.

As a summary, to conclude to the right two-class solution instead of the wrong one-class solution, BIC is reliable and most useful in all models A.8 – C.8 and with all sample sizes compared with AIC, VLMR or LMR. When the sample size is large, say over 500, instead of the BIC decision can be based on aBIC, which is more effective in finding the right two-class solution instead of the one-class solution. When  $n = 500$ , the proportion using aBIC to conclude to the wrong two-class solution is still .089, but decreases to .05 when the sample size is approximately greater than 875.

The above conclusion and results presented mean that, when SMD is 1, it is not possible to identify the right two-class solution instead of the wrong one-class solutions using AIC, BIC, aBIC, VLMR or LMR in any of the models A.8 – C.8.

When SMD is 2, it is possible to identify the true two latent classes if the sample size is large  $n = 500-1000$  by using AIC, aBIC, VLMR or LMR, but only aBIC is useful because it has acceptably low Type I error rates. When using aBIC, the proportion line is greater than .70 only in models A.8 and B.8 and the sample size needed for these models to achieve a .70, .80 or .90 proportion is 600, 800-840 or greater than 1000, respectively.

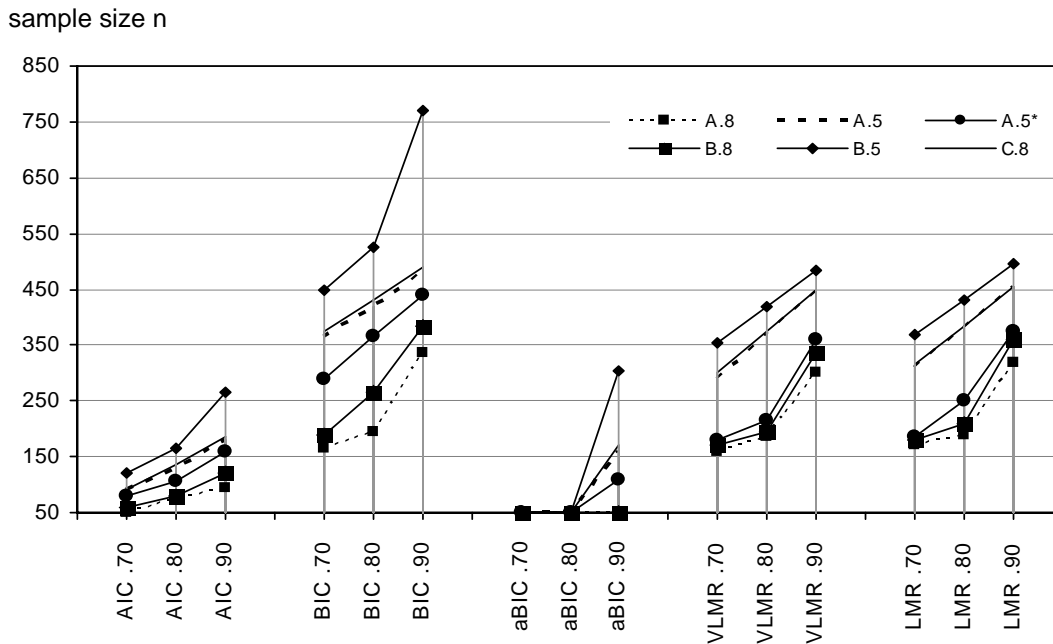


Figure 6.12. The sample size needed to achieve a .70, .80 or .90 proportion line concluding to the right two-class solution instead of to the wrong one-class solution using AIC, BIC, aBIC, VLMR and LMR in models A.8 – C.8, SMD = 3.



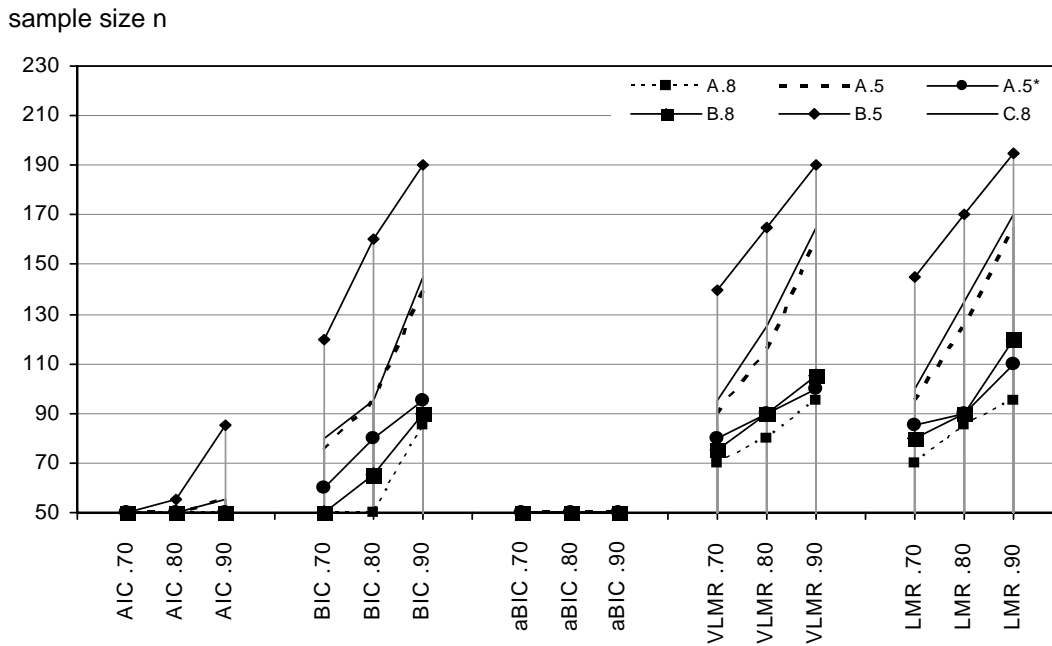


Figure 6.13. The sample size needed to achieve a .70, .80 or .90 proportion line concluding to the right two-class solution instead of to the wrong one-class solution using AIC, BIC, aBIC, VLMR and LMR in models A.8 – C.8, SMD = 4.

The sample size needed to achieve .70, .80 or .90 proportions when using AIC, BIC aBIC, VLMR and LMR are presented in the Figures 6.12 and 6.13 for all models A.8 – C.8 when SMD is 3 or 4. As can be seen from the figures, the highest sample sizes needed to achieve a .70, .80 or .90 proportion is in model B.5. After that comes the lines for model C.8 and A.5, which are very close to each other. The lowest sample size needed to achieve a .70, .80 or .90 proportion appears in model A.8 and, with a slightly larger sample size, in model B.8 using any of the criteria or tests. In model A.5\*, the sample size needed to achieve a .70, .80 or .90 proportion is clearly lower than in model A.5, and clearly greater than in model A.8 using BIC. It is noteworthy that additional measurements fully compensate with the decrease of the reliability of observed variables when using the VLMR or LMR tests, as can be seen from Figures 6.12 and 6.13.

When SMD is 3, it is possible to identify the right two-latent-classes with small sample sizes, for which purpose BIC is most appropriate. When using BIC, the proportion line .70 is achieved in models A.8, A.5\* or B.8 when the sample size is 170, 290 or 190, respectively, in models A.5 or C.8 when the sample size is 365 or 375, respectively, and in model B.5 when sample size is 450. When the sample size is 500, the proportion line for other models than model B.5 are greater than

.90 if using BIC and also in model B.5 if using aBIC. For note, when using LMR, the sample sizes needed are equal in model A.8 and in model B.8, whereas in model A.5\* the sample size needed is only 185.

When SMD is 4, the .70 proportion line is achieved with smallest sample size ( $n = 50$ ) using BIC in models A.8 and B.8. For model A.5\*, A.5, C.8 or B.5 the sample size needed to achieve the .70 proportion line is 70, 75, 80 or 120, respectively. As can be seen from Figure 6.13, when the sample size is 120, the proportion line of .70 is achieved in all models, and the .80 proportion line in all other models than model B.8, and the .90 proportion line in A.8, B.8 and A.5\*. These results describe the strong increase in proportion when the sample size increases. When SMD is 5, the proportion using BIC is greater than .90 in all models when  $n \geq 50$ .

### **6.2.2. Results of the wrong three-class solution versus the right two-class solution for model A.8**

This section presents the results concerning the proportion of decisions concluding to the wrong three-class solution instead of the right two-class solution, using model A.8. In 10000 replications, 10 random sets of starting values are produced by default by the program. These 10 sets of starting values are used in the maximum likelihood optimization with 10 iterations for each set of starting values. The ending values of two sets with the highest likelihoods are used as the starting values in the final stage of optimization, which uses the default setting for mixture analysis.

#### **6.2.2.1. Results of AIC, BIC and aBIC**

In the following Table 6.8., the proportions of wrong decisions concluding to three classes instead of to the right two classes is presented for model A.8. When using AIC, the proportion seems to be rather stable and is between .350 - .178. The proportion slightly decreases when SMD or the sample size increases, but when SMD is 5 and the sample size is 1000, the proportion is still quite high, .178.

When using BIC, the proportions are quite stable with respect to  $n$  and SMD, and the values of proportions are very small, between .001 and .066. When the sample size is 50 or 100, the proportions are between .053 and .063 or between .019 - .033, respectively. When SMD is 3, 4 or 5, the proportions are .034, .024 or .012,

respectively, when the sample size is 200 and decrease to .030, .017 or .007 when the sample size increases to 1000.

When using aBIC, the proportions are between .173 and .785 when the sample sizes are between 50 - 200, these proportions seem to be too large for making appropriate conclusions about the right number of latent classes. In turn, when the sample size is 500 or 1000, the proportions are smaller being between .060 -.116 or between .027 - .075, respectively, depending on the model.

It seems that AIC, with all SMDs and ns, and aBIC, with all SMDs when n is between 50 – 200, produce too often the wrong three-class solution. When using BIC, the proportions are very small and, especially, when the sample size is small, the proportions are near to .01 -.05 ranges, which are more suitable values for Type I error rates. Consequently, BIC seems to be the most appropriate index to evaluate the wrong three-class solution against the right two-class solution.

#### **6.2.2.2. Results of VLMR, LMR, BLRT and OLRT**

When using the VLMR or LMR test, the proportion of wrong decisions concluding to three classes instead of to the right two classes increases when the sample size or SMD increase (see Table 6.8). The proportion is lower than nominal .05 level with small sample sizes  $n = 50$  or  $100$ , and is higher than nominal .05 level when  $SMD \geq 2$  and the sample size is 500 or 1000.

When using BLRT, the proportion of wrong decisions concluding to three classes is lower than the nominal .05 level when SMD is 1. When SMD is 2 or 3, the proportion increases when sample size increases, and is .036 or .021 higher, respectively, than nominal .05 level when the sample size is 1000. When SMD is 4 or 5, the proportions are very close to nominal level.

For the VLMR, LMR, BLRT and LRT tests, the expected proportion in nominal .05 level, is .05. The VLMR and LMR tests produce low proportions when the sample size is small. When SMD is 3-5 and the sample size 500-1000, they produce high proportions, whereas BLRT produces nearly .05 proportions.

Table 6.8. The proportions of wrong decisions concluding to three classes instead of two in model A.8 when using AIC, BIC, aBIC information criteria and the VLMR, LMR, BLRT and OLRT (ordinary likelihood ratio test) tests.

n	SMD	AIC	BIC	aBIC	VLMR	LMR	BLRT	OLRT
50	1	.350	.061	.785	.022	.017	.032	.209
100	1	.295	.019	.470	.026	.020	.033	.169
200	1	.275	.007	.243	.034	.029	.032	.154
500	1	.249	.002	.084	.041	.034	.031	.133
1000	1	.236	.001	.039	.050	.044	.033	.125
50	2	.345	.054	.780	.021	.016	.028	.203
100	2	.296	.020	.474	.027	.021	.033	.170
200	2	.277	.009	.246	.041	.034	.034	.151
500	2	.279	.009	.116	.068	.057	.065	.168
1000	2	.259	.023	.075	.092	.083	.086	.157
50	3	.334	.058	.772	.026	.020	.037	.203
100	3	.290	.033	.457	.042	.035	.054	.174
200	3	.255	.034	.230	.067	.060	.077	.158
500	3	.226	.033	.087	.083	.076	.072	.130
1000	3	.198	.030	.051	.085	.078	.071	.111
50	4	.308	.066	.728	.035	.027	.051	.188
100	4	.248	.032	.404	.048	.043	.055	.139
200	4	.220	.024	.194	.065	.058	.058	.123
500	4	.197	.017	.071	.077	.069	.056	.106
1000	4	.186	.017	.038	.082	.074	.058	.101
50	5	.274	.053	.695	.039	.030	.052	.157
100	5	.217	.020	.377	.049	.041	.043	.123
200	5	.199	.012	.173	.064	.056	.046	.103
500	5	.180	.009	.060	.077	.067	.046	.094
1000	5	.178	.007	.027	.084	.075	.043	.092

### 6.2.2.3. The effect of a larger number of starting values

The former results in Table 6.8 were calculated by using 10 sets of starting values, which may not be sufficient to find a highest value of maximum likelihood. Therefore, the same data is analyzed using 500 starting values. The maximum likelihood optimization is done with 20 iterations for each starting value set. The ending values of 20 sets with the highest likelihoods are used as starting values in

the final stage of optimization, which uses default settings of iterations for mixture analysis. The current results with former results are presented in Table 6.9.

The results show that, when using AIC and adjusted BIC, overly high proportions of wrong three-class solutions increases further when starting values increase from 10 to 500, except when the sample size is high  $n = 500 - 1000$ , in which case the proportion using aBIC decreases. When using BIC, the proportion decreases and this decrease is even stronger when the sample size increases. When the sample size or SMD increases, the proportion decreases, even to .001. For the VLMR, LMR and BLRT tests, the expected proportion in nominal .05 level is .05. When using the VLMR and LMR tests, the proportions increase strongly when the sample size is large  $n = 200 - 1000$ , producing too high Type I error rate. The VLMR and LMR tests produce .09 - .19 proportions when SMD is 3-5 and the sample size 500-1000. When using BLRT, the proportions change slightly and mostly toward the nominal .05 level, the proportions being between .050 - .061.

Table 6.9. The proportions of wrong decisions concluding to three class instead of two class in model A.8 when using AIC, BIC, aBIC, VLMR, LMR, BLRT and OLRT (ordinary likelihood ratio test).

n	SMD	AIC	BIC	aBIC	VLMR	LMR	BLRT	OLRT
50	2	.345	.054	.780	.021	.016	.028	.203
		.478*	.069*	.939*	.038*	.029*	.044*	.280*
50	3	.334	.058	.772	.026	.020	.037	.203
		.464*	.066*	.933*	.045*	.034*	.050*	.272*
100	3	.290	.033	.457	.042	.035	.054	.174
		.373*	.016*	.602*	.066*	.054*	.054*	.201*
200	3	.255	.034	.230	.067	.060	.077	.158
		.318*	.005*	.279*	.106*	.091*	.060*	.171*
500	3	.226	.033	.087	.083	.076	.072	.130
		.284*	.002*	.080*	.145*	.129*	.059*	.141*
1000	3	.198	.030	.051	.085	.078	.071	.111
		.270*	.001*	.032*	.174*	.157*	.061*	.128*
50	4	.308	.066	.728	.035	.027	.051	.188
		.426*	.059*	.909*	.053*	.039*	.054*	.244*
1000	4	.186	.017	.038	.082	.074	.058	.101
		.270*	.001*	.032*	.190*	.172*	.061*	.134*

\* Same run with settings: STARTS 500 20, STITERATIONS=20

(It takes 3 - 9 days to calculate each of these results using powerful dual processing)

#### 6.2.2.4. Evaluating the information criteria to produce the wrong three-class solution versus the right two-class solution

To compare the sufficiency of penalties, an empirical distribution of  $2(\log L_3 - \log L_2)$  is shown in Table 6.10 for model A.8 with  $SMD = 3$  and the sample sizes  $n = 50, 100, 200, 500$  or  $1000$ . These results are based on the same data and 500 sets of starting values as results in Table 6.9. As can be seen from the Table 6.10, the cumulative distribution of 10000 replications for  $2(\log L_3 - \log L_2)$  values lower than 12.42, 11.36, 10.93, 10.35 and 10.03, is 95 percentage when the sample size is  $n = 50, 100, 200, 500$  or  $1000$ , respectively. These critical values are comparable with the .05 type I error rate and are 1.67 – 2.07 times larger than the penalty value of AIC, and 1.06 – 0.49 times larger than those of BIC.

Table 6.10. The empirical two times log-likelihood difference distribution and the penalties of information criteria in model A.8.

n	SMD	Penalties			Cumulative percentages and respective critical value for $2(\log L_3 - \log L_2)$					
		AIC	BIC	aBIC	70 %	80 %	90 %	95 %	99 %	99.9 %
50	2	6	11.74	2.32	7.59	8.89	10.77	12.49	16.49	23.49
50	3	6	11.74	2.32	7.52	8.71	10.70	12.42	16.75	23.03
100	3	6	13.82	4.34	6.67	7.83	9.59	11.36	15.00	20.09
200	3	6	15.90	6.39	6.17	7.36	9.14	10.93	14.43	20.50
500	3	6	18.64	9.12	5.86	6.92	8.57	10.35	13.96	19.21
1000	3	6	20.72	11.20	5.75	6.77	8.52	10.03	14.18	19.75
50	4	6	11.74	2.32	7.20	8.38	10.29	12.10	15.90	20.95
1000	4	6	20.72	11.20	5.72	6.75	8.51	10.10	13.70	19.08

The penalty value of aBIC is small when  $n = 50, 100$  or  $200$ , but is slightly lower when  $n = 500$  and slightly greater when  $n = 1000$  than 95 % cumulative value of  $2(\log L_2 - \log L_1)$ . These results indicate that the penalty value of AIC is small, producing the wrong three-class solutions in over 20 percent of cases, whereas the penalty value of BIC is high, producing in only 1- 5 percent of cases the wrong three-class solutions when the sample size is small  $n = 50$  or  $100$  and in 0.1 - 1 percent of cases when  $n = 200, 500$  or  $1000$ . The penalty of adjusted BIC is very small with small sample sizes but increases when  $n = 500$ , producing in slightly over than 5 percent of cases the wrong three-class solutions and, when  $n = 1000$ , produces in slightly less than 5 percent of cases the wrong three-class solutions.

## 6.3. Results of evaluation of parameter estimation

In this section the results of a simulation study concerning the effect of SMD and sample size, the effect of reliability, the effect of additional measurements, and the effect of model construct on parameter estimation in LGMM are described. The results concerning the evaluation of parameter estimation using the MLR estimator is presented using the next four criteria (see section 5.6)

- 1) MSE - mean square error
- 2) PB - the proportion of bias in MSE
- 3) RB - the bias of estimated standard error
- 4) 95 % coverage of parameter estimates

The results for the four criteria mentioned above are presented in the following sections 6.3.1-6.3.4.

### 6.3.1. Results of MSE

The results of the MSE for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are given in section 6.3.1.1, for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  in section 6.3.1.2, for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  in section 6.3.1.3, and for  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  in section 6.3.1.4. Section 6.3.1.5 summarizes the results of the MSE for all parameters.

In each section, the results of MSE consist of four parts. In the first part, the results of MSE in model A.8 are presented. The second part presents the effect of reliability of observed variables on the MSE, comparing model A.5 with model A.8. The third part presents the effect of additional measurement points on the MSE, comparing model A.5\* with model A.5. Finally, the fourth part outlines the effect of model construct comparing models C.8 with model A.8, model B.8 with model A.8, and model B.5 with model A.5.

### 6.3.1.1. Results of MSE for intercept parameters $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$

#### *Effect of sample size and SMD*

The MSE of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameter estimation are presented in Table 6.11. The main results suggest that these MSEs decrease when sample size or SMD increases. Because MLR estimator is consistent, the MSE decreases in all models with all SMDs as supposed. One interesting result is the large effect of SMD on the MSE. When sample size is 1000 and SMD is 1, the MSE is 79-89 larger than when SMD is 5.

When SMD is 1, the MSE in model A.8 slowly decreases when the sample size increases from  $n = 50$  to  $n = 1000$ . The MSE of  $\alpha_0^{(2)}$  decreases by only 30 percent when the sample size grows from 50 to 1000, as can be estimated from Figure 6.14. The MSE of  $\alpha_0^{(1)}$  decreases by one third when the sample size increases from about 50 to 180, from 200 to 500, or from 420 to 1000. When SMD is 2, the MSE of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  in model A.8 clearly decreases as can be seen from Figure 6.14.

When SMD is 2, the MSE of  $\alpha_0^{(2)}$  decreases by one third when the sample size increases from about 50 to 80, from 100 to 180, from 200 to 330, or from 500 to 780. For the  $\alpha_0^{(1)}$  MSE decreases by one third when the sample size increases from about 50 to 80, from 100 to 180, from 200 to 400, or from 500 to 800.

When SMD is 3, 4 or 5, both the MSE of  $\alpha_0^{(2)}$  and the MSE of  $\alpha_0^{(1)}$  decrease by half in model A.8 when the sample size increases from about  $n = 50$  to 90-100, from  $n = 100$  to 180-200, from  $n = 200$  to 400-450, or from  $n = 500$  to 1000 (see Figure 6.15).

The effect of SMD, comparing the MSE with SMD 1, 2, 3 or 4 on the MSE with SMD 5 with equal sample sizes, on the MSE in model A.8 is strong. When  $n = 50$ , the MSE of  $\alpha_0^{(1)}$  is 11.8, 4.8, 2.1 or 1.2 times larger when SMD is 1, 2, 3 or 4, respectively, than the MSE of  $\alpha_0^{(1)}$  when SMD is 5. When SMD is 1, these proportional differences increase monotonically to 78.7-fold when the sample size increases to  $n = 1000$ . When SMD is 2, these proportional differences increase to 6.8 -fold when the sample size increases to  $n = 200$  and decrease to 5.5 when the sample size increases to  $n = 1000$ . When SMD is 3 or 4, the proportional



differences decrease monotonically to 1.7 or 1.2 when the sample size increases to  $n = 1000$ .

Table 6.11. The MSE of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameter estimation in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8		A.5		A.5*		B.8		B.5		C.8	
n	SMD	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$
50	1	.4687	.4989	.7410	.7307	.7095	.6642	.4687	.4989	.7410	.7307	.5417	.5159
100	1	.3838	.4722	.6192	.6915	.6303	.6780	.3838	.4722	.6192	.6915	.4502	.4895
200	1	.2963	.4510	.5332	.6602	.5519	.6449	.2963	.4510	.5332	.6602	.3812	.4614
500	1	.1973	.3929	.3769	.5828	.3879	.5995	.1973	.3929	.3769	.5828	.2788	.4168
1000	1	.1495	.3646	.2748	.5667	.3221	.5601	.1495	.3646	.2748	.5667	.2073	.3856
50	2	.1945	.4406	.3527	.6745	.3457	.6658	.1608	.3567	.2824	.5947	.2465	.4731
100	2	.1102	.3029	.2123	.5473	.2259	.5342	.0988	.2635	.2011	.4697	.1611	.3780
200	2	.0650	.1875	.1367	.3904	.1176	.3510	.0558	.1618	.1248	.3481	.0930	.2828
500	2	.0232	.0528	.0592	.1833	.0446	.1214	.0184	.0595	.0502	.1695	.0407	.1201
1000	2	.0105	.0205	.0272	.0791	.0175	.0327	.0072	.0218	.0200	.0766	.0207	.0517
50	3	.0838	.2243	.1648	.4349	.1326	.3824	.0715	.1902	.1530	.4086	.1141	.3172
100	3	.0381	.0968	.0802	.2279	.0599	.1519	.0316	.0857	.0777	.2290	.0571	.1618
200	3	.0175	.0413	.0366	.0897	.0260	.0608	.0138	.0359	.0328	.1018	.0262	.0655
500	3	.0064	.0152	.0124	.0291	.0094	.0212	.0050	.0125	.0106	.0306	.0090	.0205
1000	3	.0032	.0074	.0061	.0138	.0046	.0105	.0025	.0061	.0049	.0138	.0044	.0099
50	4	.0486	.1185	.0891	.2240	.0694	.1724	.0435	.1020	.0881	.2237	.0618	.1628
100	4	.0228	.0532	.0405	.0960	.0321	.0746	.0201	.0459	.0380	.1006	.0280	.0691
200	4	.0111	.0253	.0190	.0442	.0155	.0357	.0099	.0215	.0173	.0424	.0133	.0313
500	4	.0043	.0099	.0073	.0169	.0060	.0136	.0012	.0084	.0065	.0155	.0051	.0120
1000	4	.0022	.0049	.0037	.0083	.0030	.0068	.0020	.0042	.0033	.0077	.0026	.0060
50	5	.0399	.0887	.0653	.1488	.0540	.1224	.0378	.0806	.0633	.1476	.0448	.1043
100	5	.0195	.0425	.0309	.0701	.0264	.0588	.0184	.0389	.0293	.0671	.0213	.0482
200	5	.0096	.0207	.0152	.0337	.0130	.0286	.0091	.0189	.0144	.0314	.0105	.0234
500	5	.0038	.0081	.0059	.0131	.0051	.0110	.0036	.0075	.0056	.0122	.0041	.0091
1000	5	.0019	.0041	.0030	.0065	.0026	.0055	.0018	.0038	.0029	.0061	.0021	.0046

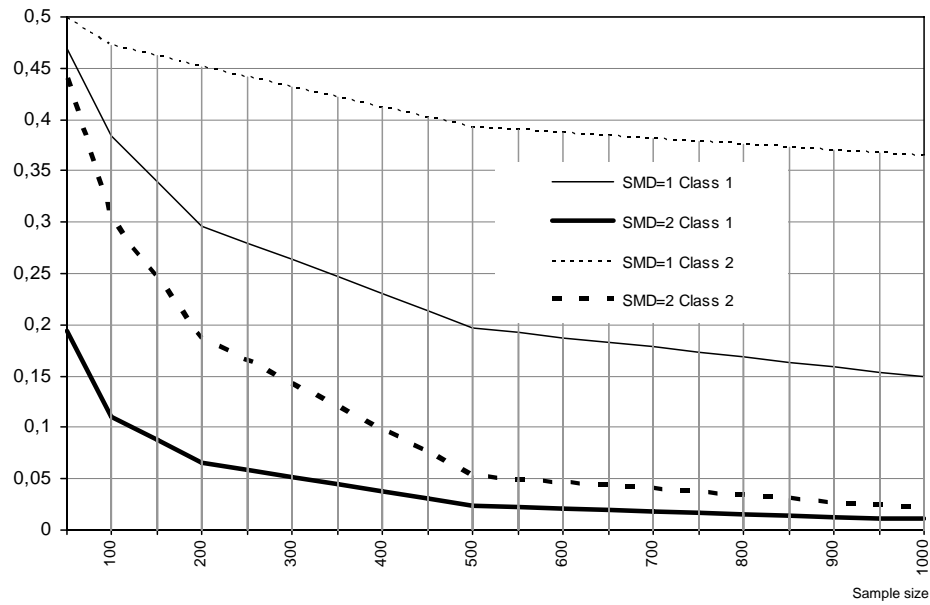


Figure 6.14. The MSE of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameter estimation in model A.8, SMD = 1 or 2.

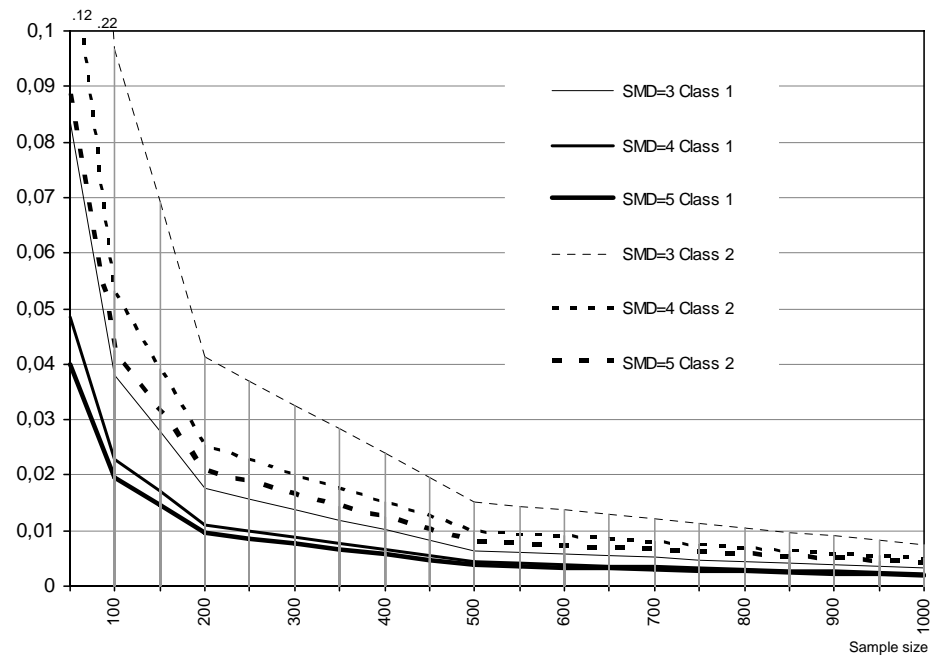


Figure 6.15. The MSE of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameter estimation in model A.8, SMD = 3, 4 or 5.

When  $n = 50$ , the MSE of  $\alpha_0^{(2)}$  is 5.62, 4.97, 2.53 or 1.34 times larger in the case where SMD is 1, 2, 3 or 4, respectively, than in the case where SMD is 5. When SMD is 1, these proportional differences increase monotonically to 88.9-fold when the sample size increases to  $n = 1000$ . When SMD is 2, these proportional differences increase to 9.1-fold when the sample size increases to  $n = 200$  and decrease then to 5.0 when the sample size increases to  $n = 1000$ . When SMD is 3 or 4, these proportional differences decrease monotonically to 1.8 or to 1.2 when the sample size increases to  $n = 1000$ .

The results presented above concerning the MSE of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameter estimation means that the convergence is slower when SMD is 1 or 2 and is faster when SMD is 3 or 4 when compared to the situation when SMD is 5.

*The effect of reliability on the MSE of estimation of intercept parameters  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

The effect of reliability on the MSE is examined by comparing the MSE of model A.5 with the MSE of model A.8 (see Table 6.11). In addition to the Figures 6.16 and 6.17, this effect is demonstrated by calculating the proportions of MSEs, that is, by dividing the MSE of model A.5 by the MSE of model A.8. The increased or decreased proportion describes how the convergence behaves as a function of sample size in the case of lower reliability than when compared with the case of higher reliability. The decreased proportion tells that the convergence is more powerful in the case of lower reliability, than in the case of higher reliability. The increased proportion tells the conversed pattern of results.

When the reliability of observed variables decreases from .80 to .50 (model A.8 vs. model A.5), the MSE of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  strongly increase (see Table 6.11). The largest proportions in the MSE are seen when SMD is 2 or 3. When SMD is 2, the largest proportion for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters are 2.6 or 3.9, respectively. When SMD is 3 and  $n = 200$ , the MSE in model A.5, is 2.1 – 2.2 times larger than in model A.8. When SMD is 5, the MSE in model A.5, is 1.6 times larger than in model A.8. These results mean that the sample size has to be two times larger in model A.5 than in model A.8, in order to get approximately equal MSE of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$ .

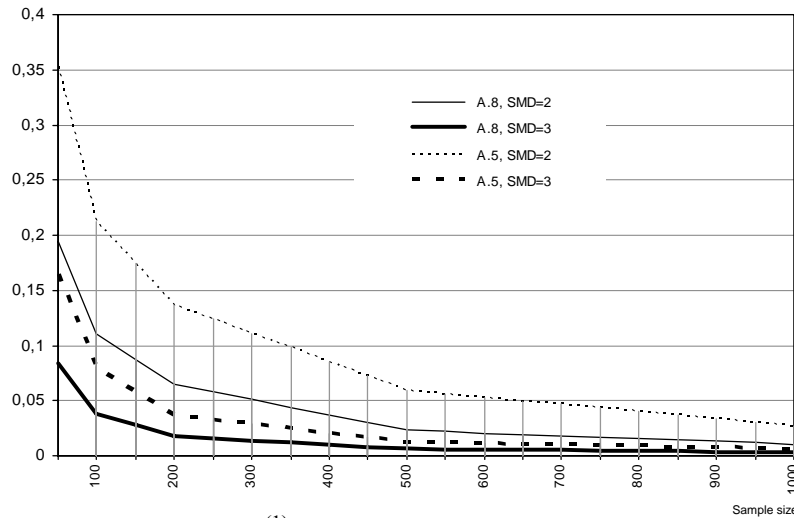


Figure 6.16. The MSE of  $\alpha_0^{(1)}$  parameter estimation in models A.8 and A.5, SMD = 2 or 3.

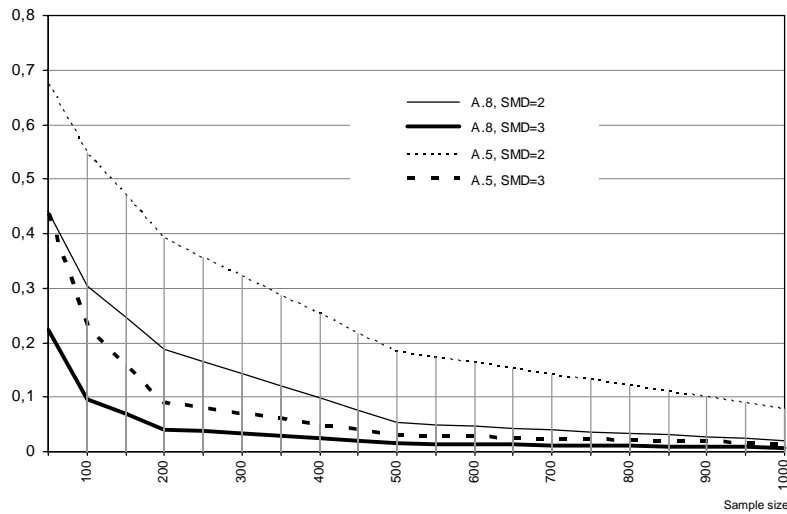


Figure 6.17. The MSE of  $\alpha_0^{(2)}$  parameter estimation in models A.8 and A.5, SMD = 2 or 3.

The pattern of clearly decreasing MSE with different SMD = 2, 3, 4 or 5 is very similar to model A.5 and model A.8, as can be seen, for example, from Figures 6.16 and 6.17. When  $n = 50$ , the MSE of  $\alpha_0^{(1)}$  in model A.5, is 1.6, 1.8, 2.0, 1.8 or 1.6 times larger with SMD = 1, 2, 3, 4 or 5, respectively, than the MSE of  $\alpha_0^{(1)}$  in model A.8 (see Table 6.11). When SMD is 1 or 2, these proportions increase monotonically to 1.8 or 2.6 –fold when the sample size increases to  $n = 1000$ . When SMD is 3, these proportions increase to 2.1 –fold when the sample size

increases to  $n = 100$ , and then decrease to 1.9 when the sample size increases to  $n = 1000$ . When SMD is 4, these proportions decrease monotonically to 1.7 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion is very stable, being 1.6.

When  $n = 50$ , the MSE of  $\alpha_0^{(2)}$  in model A.5, is 1.5, 1.5, 1.9, 1.9 or 1.7 times larger when SMD is 1, 2, 3, 4 or 5, respectively, than the MSE of  $\alpha_0^{(2)}$  in model A.8. When SMD is 1 or 2, these proportions increase monotonically to 1.6 or 3.9 –fold when the sample size increases to  $n = 1000$ . When SMD is 3, these proportions increase to 2.4 –fold when the sample size increases to  $n = 100$  and decrease then to 1.9 when the sample size increases to  $n = 1000$ . When SMD is 4, these proportions decrease monotonically to 1.7 when the sample size increases to  $n = 1000$ . When SMD is 5, proportion is very stable, being 1.6.

The results mean that the convergence of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameter estimates is more rapid in model A.8 than in model A.5 when SMD is 1 or 2. When SMD is 3, the convergence of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameter estimates is first slower in model A.5 than in model A.8, but after the sample size is 100 or larger, it is more rapid in model A.5 than in model A.8. When SMD is 4, the convergence for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameter estimates is more rapid in model A.5 than in model A.8. When SMD is 5, the convergence is equal and the effect of reliability stabilizes, the MSE being 1.6 times larger in model A.5 than in model A.8.

*The effect of additional measurements on the MSE of estimation of intercept parameters  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

The effect of additional measurement points on the MSE is examined by comparing the MSE in model A.5\* with the MSE in model A.5 (see Table 6.11). Additional measurement points can compensate the lack of reliability and, therefore, the MSE of parameters in model A.5\* are mainly located between the MSE of model A.8 and the MSE of model A.5., as can be seen, for example, from Figures 6.18 and 6.19.

The effect of additional measurement points is described in the following percentage:

$$Percentage = \frac{MSE(A.5^*) - MSE(A.5)}{MSE(A.8) - MSE(A.5)} * 100 \quad (6.1)$$

By this way, the distance of the MSE in A.5\* from the MSE in A.5 is described as a percentage of the distance of the MSE in A.8 from the MSE in A.5. This percentage is 0 or 100 if the MSE in model A.5\* is equal with the MSE in model A.5 or equal with the MSE in A.8, respectively. The percentage can be negative, meaning that the MSE in model A.5\* is greater than in model A.5 and greater than 100, meaning that the MSE in model A.5\* is lower than in model A.8.

For  $\alpha_0^{(1)}$  parameter, the MSE in model A.5\* is almost equal with that in model A.5 when SMD is 1. When the sample size increases from  $n = 50$  to  $n = 1000$ , the distance between the MSE in model A.5\* and the MSE in model A.5 increases for  $\alpha_0^{(1)}$  parameter from 4 % to 58 % when SMD is 2, or from 40 % to 52 % when SMD is 3, and decreases from 49 % to 47% when SMD is 4 or from 44 % to 36 % when SMD is 5.

For  $\alpha_0^{(2)}$  parameter, the MSE in model A.5\* is almost equal to that in model A.5 when SMD is 1. When the sample size increases from  $n = 50$  to  $n = 1000$ , the distance between the MSE in model A.5\* and the MSE in model A.5 increases for  $\alpha_0^{(2)}$  parameter from 4 % to 79 % when SMD is 2, or from 25 % to 52 % when SMD is 3, and decreases when SMD is 4 or 5 from 49 % to 44% or from 44 % to 42 %, respectively.

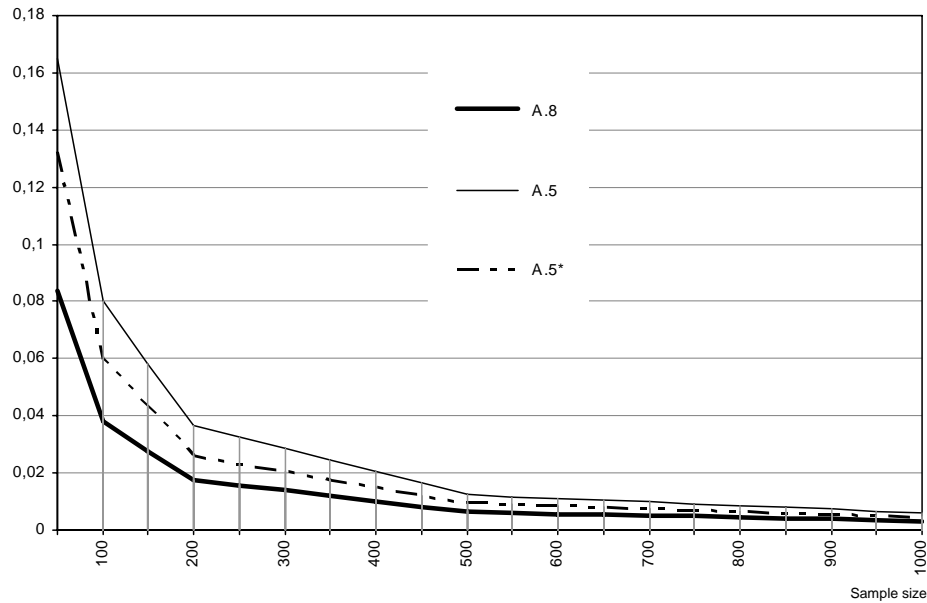


Figure 6.18. The MSE of  $\alpha_0^{(1)}$  parameter estimation as a function of sample size in models A.8, A.5 and A.5\*, SMD = 3.

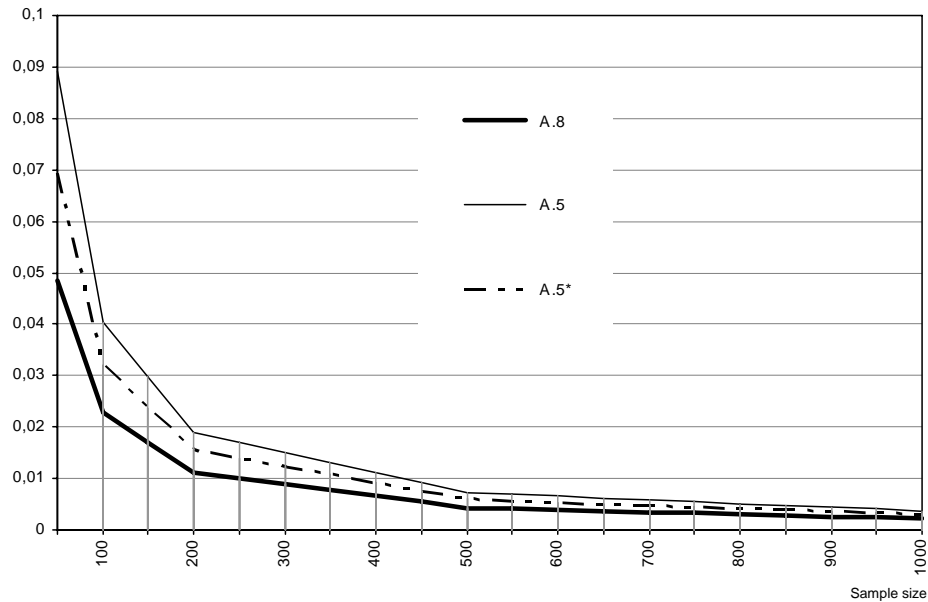


Figure 6.19. The MSE of  $\alpha_0^{(2)}$  parameter estimation as a function of sample size in models A.8, A.5 and A.5\*, SMD = 3.

When SMD is 1, the MSE decreases slowly in model A.5\*. When SMD is 2, the MSE decreases for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters in model A.5\* more rapidly than in model A.5, and when SMD is 4 or 5, the MSE decreases almost equally in both models.

*The effect of model construct on the MSE of estimation of intercept parameters  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

The construct of the model has a clear effect on the MSE of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters and is examined through comparing model C.8 with model A.8, model B.8 with model A.8, and model B.5 with model A.5 (see Table 6.11). The effect of the construct is described by using the proportion of the MSEs. For example, the proportion comparing model B.8 and A.8, is calculated by dividing the MSE of model B.8 by the MSE of model A.8.

*Model C.8 vs. model A.8 -  $\alpha_0^{(1)}$*

The estimation of parameter  $\alpha_0^{(1)}$  is more effective in model A.8 than in model C.8, and when  $n = 50$ , the MSE in model C.8 is 1.16, 1.27, 1.36, 1.27 and 1.12 times larger than the MSE in model A.8 when SMD is 1, 2, 3, 4 or 5, respectively. The

proportion increases to 1.97 when SMD is 1 and to 1.97 when SMD is 2 when the sample size increases to  $n = 1000$ . If SMD is 3, the proportion first increases from 1.36 to 1.50 when the sample size increases from 50 to 100 and, after that, decreases to 1.38 when the sample size increases to 1000. When SMD is 4, the proportion decreases from 1.27 to 1.18 when the sample size increases from  $n = 50$  to  $n = 1000$ . When SMD is 5, the proportion is very stable, being 1.11.

*Model B.8 vs. model A.8 -  $\alpha_0^{(1)}$*

For parameter  $\alpha_0^{(1)}$ , the estimation is more effective in model B.8 than in model A.8. When  $n = 50$ , the MSE in model B.8 is 0.83, 0.85, .90 or .95 times the MSE in model A.8 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, the proportion of the MSEs first increases to .90 when the sample size increases to 100 and decreases after that to .69 when the sample size increases to 1000. When SMD is 3, the proportion decreases to .78 when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportion is very stable with different sample sizes and ranges between .89 - .91 or between .94 - .95, respectively.

*Model B.5 vs. model A.5 -  $\alpha_0^{(1)}$*

Comparing the MSE in model B.5 with the MSE in model A.5, the effect is similar as in the case of higher reliability. The estimation of parameter  $\alpha_0^{(1)}$  is more effective in model B.5 than in model A.5. When  $n = 50$ , the MSE in model B.8 is 0.80, 0.93, .99 or .97 times the MSE in model A.8 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, the proportion first increases to .95 when the sample size increases to  $n = 100$ , and decreases after that to .74 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion first increases from .97 when the sample size increases to  $n = 100$ , and decreases after that to .80 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportion decreases to .89 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion is very stable and decreases to .95 when the sample size increases from  $n = 50$  to  $n = 200$  and increases after that to .97 when the sample size increases to  $n = 1000$ .

*Model C.8 vs. model A.8 -  $\alpha_0^{(2)}$*

The estimation of parameter  $\alpha_0^{(2)}$  is more effective in model A.8 than in model C.8, and when the sample size is 50, the MSE in model C.8 is 1.03, 1.07, 1.41, 1.37 and 1.18 times larger than the MSE in model A.8 when SMD is 1, 2, 3, 4 or 5, respectively. When SMD is 1, the proportion increases to 1.06 when the sample size increases to  $n = 1000$ . When SMD is 2, the proportion increases to 2.52 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion increases to 1.67 when the sample size increases to 100 and, after that, decreases to 1.34 when



the sample size increases to 1000. When SMD is 4, the proportion decreases to 1.22 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion decreases to 1.13 when the sample size increases to  $n = 100$  and is very stable, 1.12, when the sample size is greater or equal to  $n = 200$ .

*Model B.8 vs. model A.8 -  $\alpha_0^{(2)}$*

For parameter  $\alpha_0^{(2)}$ , the estimation is more effective in model B.8 than in model A.8. When the sample size is 50, the MSE in model B.8 is .81, .85, .86 or .91 times the MSE in model A.8 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, the proportion increases to 1.06 when the sample size increases to  $n = 1000$ . When SMD is 3, 4 or 5, the proportion is very stable with all sample sizes and, when the sample size is  $n = 1000$ , the proportion is .82, .86 or .93, respectively.

*Model B.5 vs. model A.5 -  $\alpha_0^{(2)}$*

For parameter  $\alpha_0^{(2)}$ , the estimation is slightly more effective in model B.5 than in model A.5 when SMD is 4 or 5 and  $n = 200, 500$  or  $1000$ . The proportion is between 0.88-1.14 when comparing the MSE in model B.5 with the MSE in model A.5. When SMD is 2, the proportion increases from .88 to .97 when the sample size increases from  $n = 50$  to  $n = 1000$ . When SMD is 3, the proportion increases from .94 to 1.13 when the sample size increases from  $n = 50$  to  $n = 200$ , and decreases after that to 1.0 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportion increases from 1 to 1.05 when the sample size increases from  $n = 50$  to  $n = 100$  and decreases to .93 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion decreases from .99 to .94 when the sample size increases from  $n = 50$  to  $n = 1000$ .

**6.3.1.2. Results of MSE for slope parameters  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$**

The results of the MSE for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimation are presented in Table 6.12. The running order is the same as for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters in section 6.3.1.1. The results suggest that the value of MSE depends on both SMD and sample size. The effect of sample size with different SMDs is described in Figures 6.20 and 6.21.

When SMD is 1, the MSE in model A.8 slowly decreases when the sample size increases from  $n = 50$  to  $n = 1000$ . For  $\alpha_1^{(1)}$  parameter, the MSE decreases by one third when the sample size grows from about 50 to 320, from 200 to 750, or from

500 to 1000, whereas the MSE of  $\alpha_1^{(2)}$  decreases only by 35 percent when the sample size grows from 50 to 1000. When SMD is 2, the MSE of  $\alpha_1^{(1)}$  parameter decreases by one third when the sample size grows from about 50 to 95, from 100 to 180, from 200 to 340, or from 500 to 750 (see Figure 6.20). For  $\alpha_1^{(2)}$  parameter, the MSE decreases by one third when the sample size grows from about 50 to 120, from 100 to 200, from 200 to 360, or from 500 to 750. When SMD is 3, 4 or 5, both the MSE of  $\alpha_1^{(1)}$  and the MSE of  $\alpha_1^{(2)}$  decrease by half when the sample size grows from about 50 to 90-100, from 100 to 180-200, from 200 to 400-450, or from 500 to 1000 (see figure 6.21), which are similar to results found for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters in model A.8.

The effect of SMD, comparing the MSE with SMD 1, 2, 3 or 4 to the MSE with SMD 5 with equal sample sizes, on the MSE in model A.8 is strong. When  $n = 50$ , the MSE of  $\alpha_1^{(1)}$  is 7.1, 3.9, 1.7 or 1.1 times larger when SMD is 1, 2, 3 or 4, respectively, than the MSE of  $\alpha_1^{(1)}$  when SMD is 5. When SMD is 1, these proportional differences increase monotonically to 70.3-fold when the sample size increases to  $n = 1000$ . When SMD is 2, these proportional differences increase to 6.6-fold when the sample size increases to  $n = 200$ , and decrease then to 4.0 when the sample size increases to  $n = 1000$ . When SMD is 3 or 4, the proportional difference decreases to 1.25 or 1.0, respectively, when the sample size increases to  $n = 1000$ .

When  $n = 50$ , the MSE of  $\alpha_1^{(2)}$  is 5.0, 4.5, 2.1 or 1.2 times larger when SMD is 1,2,3 or 4, respectively, than the MSE of  $\alpha_1^{(2)}$  when SMD is 5. When SMD is 1, these proportional differences increase monotonically to 69.1-fold when the sample size increases to  $n = 1000$ . When SMD is 2, these proportional differences increase to 8.8-fold when the sample size increases to  $n = 200$  and decrease then to 6.2 when the sample size increases to  $n = 1000$ . When SMD is 3 or 4, the proportional difference decreases to 1.6 or 1.1, respectively, when the sample size increases to  $n = 1000$ .

These results for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimates in model A.8 mean that the convergence is slower when SMD is 1, faster when SMD is 3, and equal when SMD is 4, compared with the convergence when SMD is 5.

Table 6.12. The MSE of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimation in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8		A.5		A.5*		B.8		B.5		C.8	
n	SMD	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$
50	1	.0653	.0956	.1328	.1781	.5417	.5159	.0653	.0956	.1328	.1781	.0750	.1001
100	1	.0564	.0855	.1149	.1592	.4502	.4895	.0564	.0855	.1149	.1592	.0651	.0919
200	1	.0480	.0778	.0979	.1511	.3812	.4614	.0480	.0778	.0979	.1511	.0548	.0867
500	1	.0366	.0725	.0784	.1367	.2788	.4168	.0366	.0725	.0784	.1367	.0474	.0752
1000	1	.0281	.0622	.0593	.1214	.2073	.3856	.0281	.0622	.0593	.1214	.0390	.0737
50	2	.0358	.0848	.0869	.1700	.2465	.4731	.0431	.0989	.0933	.1935	.0490	.0982
100	2	.0234	.0602	.0588	.1367	.1611	.3780	.0244	.0689	.0624	.1533	.0351	.0817
200	2	.0145	.0406	.0390	.1020	.0930	.2828	.0140	.0445	.0398	.1144	.0205	.0627
500	2	.0039	.0143	.0148	.0543	.0407	.1201	.0055	.0146	.0160	.0536	.0092	.0310
1000	2	.0016	.0056	.0062	.0235	.0207	.0517	.0022	.0051	.0069	.0221	.0038	.0141
50	3	.0156	.0397	.0403	.1050	.1141	.3172	.0202	.0556	.0526	.1427	.0229	.0614
100	3	.0071	.0187	.0203	.0592	.0571	.1618	.0096	.0246	.0276	.0798	.0119	.0319
200	3	.0032	.0081	.0083	.0246	.0262	.0655	.0043	.0101	.0123	.0338	.0049	.0143
500	3	.0011	.0028	.0027	.0074	.0090	.0205	.0016	.0037	.0041	.0108	.0015	.0044
1000	3	.0005	.0014	.0013	.0035	.0044	.0099	.0008	.0019	.0019	.0047	.0007	.0020
50	4	.0102	.0227	.0221	.0545	.0618	.1628	.0172	.0304	.0312	.0831	.0129	.0312
100	4	.0048	.0105	.0102	.0244	.0280	.0691	.0058	.0134	.0141	.0353	.0058	.0137
200	4	.0024	.0051	.0048	.0111	.0133	.0313	.0029	.0063	.0065	.0153	.0028	.0064
500	4	.0009	.0020	.0018	.0042	.0051	.0120	.0011	.0025	.0025	.0058	.0011	.0025
1000	4	.0004	.0010	.0009	.0021	.0026	.0060	.0006	.0013	.0012	.0029	.0005	.0012
50	5	.0092	.0190	.0182	.0392	.0448	.1043	.0102	.0223	.0227	.0542	.0104	.0222
100	5	.0045	.0093	.0087	.0186	.0213	.0482	.0049	.0107	.0105	.0240	.0049	.0105
200	5	.0022	.0046	.0043	.0091	.0105	.0234	.0024	.0052	.0051	.0115	.0025	.0052
500	5	.0009	.0018	.0017	.0036	.0041	.0091	.0010	.0021	.0020	.0045	.0010	.0021
1000	5	.0004	.0009	.0008	.0018	.0021	.0046	.0005	.0010	.0010	.0023	.0005	.0010

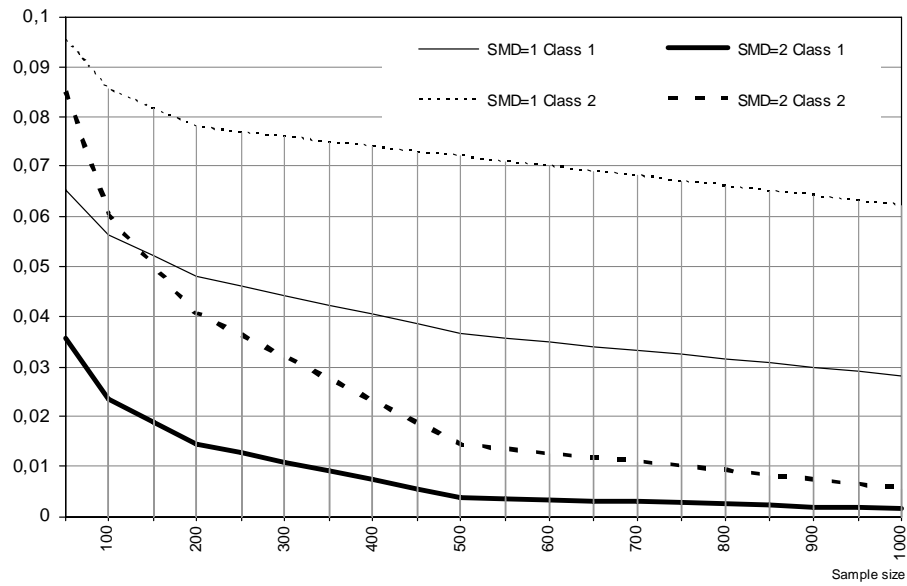


Figure 6.20. The MSE of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimation as a function of sample size, SMD=1 or 2.

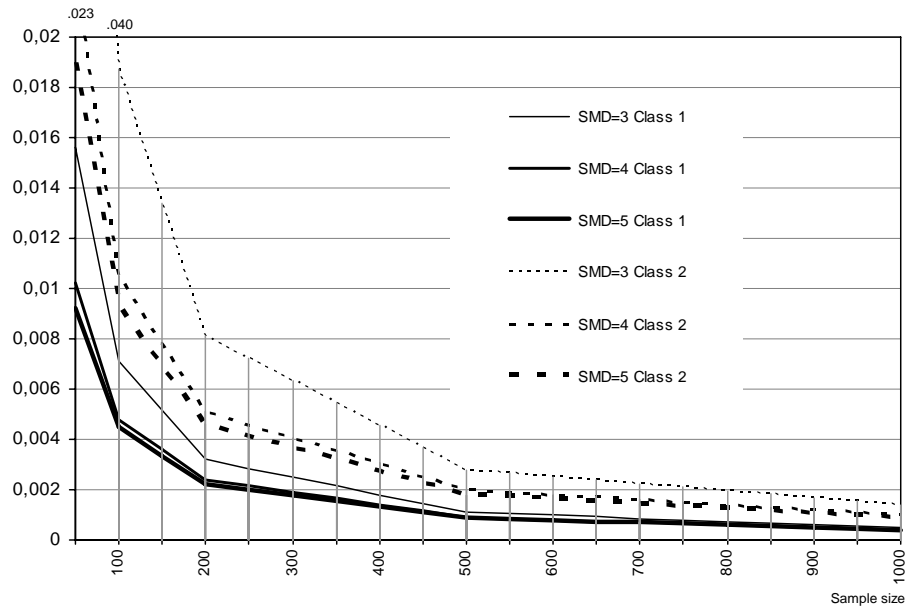


Figure 6.21. The MSE of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimation as a function of sample size, SMD=3, 4 or 5.

*The effect of reliability on the MSE of slope parameter  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  estimation*

The effect of reliability on the MSE of slope parameter estimation is examined comparing the MSE of model A.5 with the MSE of model A.8 using proportions of MSEs (see Table 6.12). These proportions are calculated by dividing the MSEs of slope parameter estimation in model A.5 by the MSEs of slope parameter estimation in model A.8.

When the reliability of observed variables decreases from .80 to .50 (model A.8 vs. model A.5), the MSE of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimation strongly increases. The largest proportions are seen when SMD is 2 or 3. When SMD is 2, the largest proportion for  $\alpha_1^{(1)}$  or  $\alpha_1^{(2)}$  parameter is 3.9 or 4.2, respectively. When SMD is 3 and  $n = 100$ , the MSE for  $\alpha_1^{(1)}$  or  $\alpha_1^{(2)}$  parameters is 2.9 or 3.2 times larger, respectively, in model A.5 than in model A.8. When SMD is 5, the MSE in model A.5, is 2 times larger than in model A.8. These results mean that the sample size has to be at least two times larger in model A.5 than in model A.8 for the result on the MSE of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimation in model A.5 to be near the MSE in model A.8.

When  $n = 50$ , the MSE of  $\alpha_1^{(1)}$  in model A.5 is 2.0, 2.4, 2.6, 2.2 or 2.0 times larger when SMD is 1,2,3,4 or 5, respectively, than the MSE of  $\alpha_1^{(1)}$  in model A.8. When SMD is 1 or 2, these proportions increase monotonically to 2.1 or 3.9 –fold when the sample size increases to  $n = 1000$ . When SMD is 3, these proportional differences increase to 2.9 –fold when the sample size increases to  $n = 100$ , and decrease then to 2.6 when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportion is very stable being about 2.0.

When  $n = 50$ , the MSE of  $\alpha_1^{(2)}$  in model A.5 is 1.9, 2.0, 2.6, 2.4 or 2.1 times larger when SMD is 1, 2, 3, 4 or 5, respectively, than the MSE of  $\alpha_1^{(2)}$  in model A.8. When SMD is 1, the proportion of MSEs is very stable, whereas when SMD is 2, this proportion increases monotonically and is 4.2 –fold when the sample size increases to  $n = 1000$ . When SMD is 3, this proportion increases to 3.2 –fold when the sample size increases to  $n = 100$  and decreases to 2.5 -fold when the sample size increases to  $n = 1000$ . When SMD is 4, this proportion decreases monotonically to 2.1 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion is very stable being 2.0.

The results mean that when SMD is 2, the convergence of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimates is more rapid in model A.8 than in model A.5. When SMD is 3, the convergence of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  is first slower in model A.5 than in model A.8, but after the sample size increases to  $n = 100$  or over, it is more rapid in model A.5 than in model A.8. When SMD is 4, the convergence of  $\alpha_0^{(2)}$  is more rapid in model A.5 than in model A.8, whereas that of  $\alpha_0^{(1)}$  is equal in both models. When SMD is 5, the convergences are equal in both models. In this case, the effect of reliability stabilizes and the MSE in model A.5 is 2.0 times larger than in model A.8.

*The effect of additional measurements on the MSE of slope parameter  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  estimation*

The effect of additional measurement points on the MSE of slope parameter estimates is examined by comparing the MSE in model A.5\* with the MSE in model A.5 (see Table 6.12). As before for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$ , the effect of additional measurement points is described as percentages of distance.

The MSE of  $\alpha_1^{(1)}$  parameter is lower in model A.5\* than in model A.5 with all sample sizes and SMDs, except when SMD is 1 and  $n = 1000$ . When SMD is 2, the distance between the MSE of  $\alpha_1^{(1)}$  in model A.5\* and the MSE of  $\alpha_1^{(1)}$  in model A.5 increases from 10 % to 109 % when the sample size increases from  $n = 50$  to  $n = 1000$ . When SMD is 3, the distance first increases from 54 % to 80 % when the sample size increases from  $n = 50$  to  $n = 100$  and, after that, decreases to 69 % when the sample size increases to  $n = 500$  and again increases to 113 % when the sample size increases to  $n = 1000$ . When SMD is 4, the distance decreases from 54 % to 40 % when the sample size increases from  $n = 50$  to  $n = 1000$  and when SMD is 5, the distance decreases in this case from 47 % to 25%.

The parameter estimation of  $\alpha_1^{(2)}$  is more effective in model A.5\* than in model A.5. The MSE of  $\alpha_1^{(2)}$  in model A.5\* is lower than MSE of  $\alpha_1^{(2)}$  in model A.5 and, when the sample size increases from  $n = 50$  to  $n = 1000$ , the percentage of distance increases from 8 % to 49 % when SMD is 1, from 45 % to 105 % when SMD is 2, and from 52 % to 81 % when SMD is 3. When SMD is 4 or 5, the distance slightly varies and is 55 % or 44 %, respectively, when the sample size is  $n = 1000$ .

When SMD is 1, the convergence of  $\alpha_1^{(1)}$  or  $\alpha_1^{(2)}$  is slow in model A.5\*. When SMD is 2 or 3, the convergence of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  is more rapid in model A.5\* than in model A.5. When SMD is 3 and  $n = 1000$ , the MSE in model A.5\* is even lower than in model A.8. When SMD is 4 or 5, the convergence of  $\alpha_1^{(2)}$  parameter is slower in model A.5\* than in model A.5, whereas the convergence of  $\alpha_1^{(1)}$  parameter is equal in models A.5 and A.5\*

*The effect of model construct on the MSE of intercept parameter  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  estimation*

The construct of the model has a clear effect on the MSE of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimation. This effect is examined comparing model C.8 with model A.8, model B.8 with model A.8, and model B.5 with model A.5 (see Table 6.12). As before, the effect of construct is described using the proportion of the MSEs for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters. For example, the proportion comparing model B.8 and model A.8 is calculated by dividing the MSE of model B.8 by the MSE of model A.8.

The estimation of parameter  $\alpha_1^{(1)}$  is more effective in model A.8 than in model C.8, and, when  $n = 50$ , the MSE in model C.8 is 1.15, 1.37, 1.47, 1.26 and 1.13 times larger than the MSE in model A.8 when SMD is 1, 2, 3, 4 or 5, respectively. When the sample size increase to  $n = 1000$ , the proportion increases to 1.39 when SMD is 1 and to 2.38 when SMD is 2. When SMD is 3, the proportion first increases to 1.69 when the sample size increases to 100 and, after that, decreases to 1.40 when the sample size increases to 1000. When SMD is 4 or 5, the proportion slightly varies and is 1.25 when  $n = 1000$ .

The estimation of parameter  $\alpha_1^{(1)}$  is more effective in model A.8 than in model B.8, and when  $n = 50$ , the MSE in model B.8 is 1.20, 1.29, 1.69 and 1.11 times larger than the MSE in model A.8 when SMD is 2,3,4 or 5, respectively. When SMD is 2, the proportion first decreases to .97 when the sample size increases to  $n = 200$  and, after that, increases to 1.38 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion increases to 1.60 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportion decreases first to 1.21 when the sample size increases to  $n = 200$  and, after that, increases to 1.5 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion increases to 1.25 when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\alpha_1^{(1)}$  is more effective in model A.5 than in model B.5, and when  $n = 50$ , the MSE in model B.5 is 1.07, 1.31, 1.41 and 1.25 times larger than the MSE in model A.5 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, the proportion slightly varies and is 1.11 when  $n = 1000$ . When SMD is 3, the proportion first increases to 1.52 when the sample size increases to  $n = 500$  and, after that, decreases to 1.46 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportion decreases to 1.33 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion is very stable and is 1.25 when  $n = 1000$ .

The estimation of parameter  $\alpha_1^{(2)}$  is more effective in model A.8 than in model C.8, and when the sample size is 50, the MSE in model C.8 is 1.05, 1.16, 1.55, 1.37 or 1.17 times larger than the MSE in model A.8 when SMD is 1, 2, 3, 4 or 5, respectively. When SMD is 1, the proportion increases to 1.18 when the sample size increases to  $n = 1000$ . When SMD is 2, the proportion increases to 2.52 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion first increases to 1.77 when the sample size increases to 200 and, after that, decreases to 1.43 when the sample size increases to 1000. When SMD is 4 or 5, the proportion decreases to 1.20 or 1.11, respectively.

The estimation of parameter  $\alpha_1^{(2)}$  is more effective in model A.8 than in model B.8, and when  $n = 50$ , the MSE in model B.8 is 1.16, 1.40, 1.34 or 1.17 times larger than the MSE in model A.8 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, the proportion decreases to 0.91 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion first decreases to 1.25 when the sample size increases to  $n = 200$  and, after that, increases to 1.36 when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportion slightly varies and is 1.30 or 1.11, respectively, when  $n = 1000$ .

The estimation of parameter  $\alpha_1^{(2)}$  is more effective in model A.5 than in model B.5, and when  $n = 50$ , the MSE in model B.5 is 1.14, 1.36, 1.52 or 1.38 times larger than the MSE in model A.5 when SMD is 2,3,4 or 5, respectively. When SMD is 2, the proportion decreases to 0.94 when the sample size increases to  $n = 1000$ . When SMD is 3, 4 or 5, the proportion slightly varies and is 1.34, 1.38 or 1.28, respectively, when  $n = 1000$ .



### 6.3.1.3. Results of MSE of variances $\psi_{00}$ , $\psi_{11}$ and covariance $\psi_{01}$ parameter estimation

The results of MSE for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameter estimation are presented in Table 6.13. The running order is the same as for the intercept and slope parameters in sections 6.3.1.1 and 6.3.1.2. The results suggest that the value of the MSE depends on both SMD and sample size. The effect of sample size in model A.8 with different SMD is described in Figures 6.22, 6.23 and 6.24.

When SMD is 1, the MSE of  $\psi_{00}$  in model A.8 clearly decreases when the sample size increases from  $n = 50$  to  $n = 1000$ . For  $\psi_{00}$  parameter, the MSE decreases by one third when the sample size grows from about 50 to 95, from 100 to 180, from 200 to 400, or from 500 to 930.

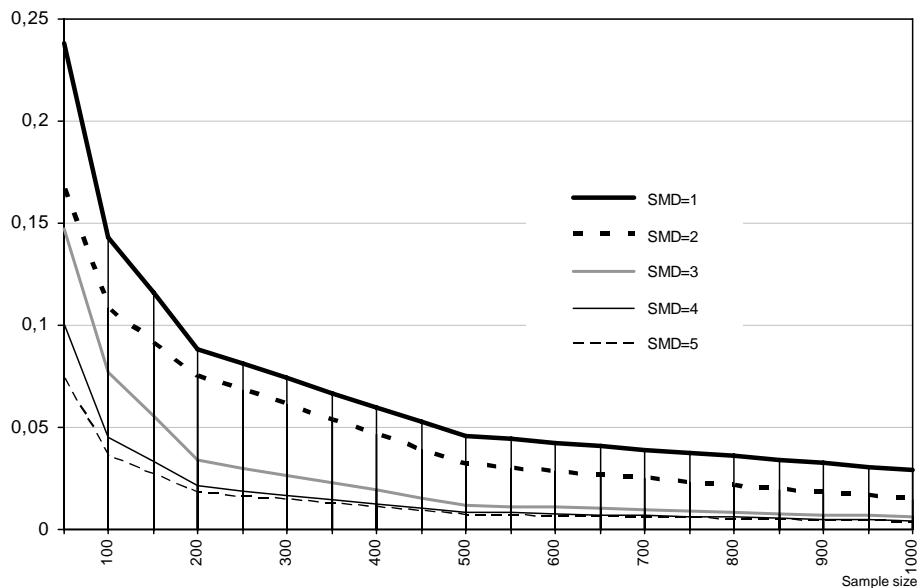


Figure 6.22. The MSE of the  $\psi_{00}$  parameter estimation as a function of sample size, SMD=1, 2, 3, 4 and 5.

When SMD is 2, the MSE of  $\psi_{00}$  decreases by one third when the sample size grows from about 50 to 95, from 100 to 220, from 200 to 380 or from 500 to 780 (see Figure 6.22). When SMD is 3, the MSE of  $\psi_{00}$  decreases by half when the sample size grows from about 50 to 110, from 100 to 180, from 200 to 420 or from 500 to 1000. When SMD is 4, the MSE of  $\psi_{00}$  decreases by half when the sample

size grows from about 50 to 95, from 100 to 200, from 200 to 430 or from 500 to 1000. When SMD is 5, the MSE of  $\psi_{00}$  decreases by half when the sample size grows from about 50 to 100, from 100 to 230, from 200 to 480 or from 500 to over 1000.

Table 6.13. The MSE of  $\psi_{00}, \psi_{11}, \psi_{01}$  parameter estimation in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model	A.8	A.5	A.5*	B.8	B.5	C.8
n SMD	$\psi_{00}$ $\psi_{11}$ $\psi_{01}$	$\psi_{00}$ $\psi_{11}$ $\psi_{01}$	$\psi_{00}$ $\psi_{11}$ $\psi_{01}$	$\psi_{00}$ $\psi_{11}$ $\psi_{01}$	$\psi_{00}$ $\psi_{11}$ $\psi_{01}$	$\psi_{00}$ $\psi_{11}$ $\psi_{01}$
50 1	.2380.0070.0196	.5869.0299.0718	.2911.0169 -	.2380.0070.0196	.5869.0299.0718	.2612.0084.0218
100 1	.1433.0045.0140	.3559.0190.0468	.1839.0097 -	.1433.0045.0140	.3559.0190.0468	.1612.0054.0153
200 1	.0880.0029.0097	.2216.0116.0314	.1140.0053 -	.0880.0029.0097	.2216.0116.0314	.1002.0035.0104
500 1	.0459.0016.0059	.1168.0064.0179	.0628.0025 -	.0459.0016.0059	.1168.0064.0179	.0524.0020.0064
1000 1	.0291.0010.0040	.0718.0040.0118	.0421.0013 -	.0291.0010.0040	.0718.0040.0118	.0322.0013.0044
50 2	.1668.0060.0243	.4543.0280.0841	.2439.0138 -	.1348.0097.0181	.3554.0378.0628	.1807.0079.0267
100 2	.1082.0033.0162	.2714.0156.0520	.1628.0068 -	.0797.0060.0122	.2067.0224.0389	.1127.0045.0179
200 2	.0752.0016.0102	.1761.0083.0331	.1050.0031 -	.0454.0038.0074	.1189.0136.0234	.0770.0024.0117
500 2	.0319.0005.0045	.0935.0031.0160	.0491.0009 -	.0194.0017.0029	.0540.0066.0102	.0418.0009.0060
1000 2	.0146.0002.0021	.0512.0013.0084	.0220.0004 -	.0090.0008.0011	.0267.0035.0044	.0229.0004.0033
50 3	.1473.0046.0220	.4316.0238.0839	.2202.0101 -	.0945.0091.0185	.2846.0378.0652	.1741.0063.0267
100 3	.0771.0022.0113	.2473.0114.0458	.1047.0043 -	.0479.0048.0094	.1474.0220.0353	.1028.0031.0155
200 3	.0337.0010.0052	.1219.0051.0226	.0466.0020 -	.0224.0022.0041	.0703.0113.0167	.0489.0014.0077
500 3	.0120.0004.0019	.0419.0019.0081	.0163.0008 -	.0085.0008.0015	.0263.0041.0058	.0165.0005.0028
1000 3	.0060.0002.0009	.0200.0010.0039	.0082.0004 -	.0042.0004.0007	.0126.0020.0027	.0081.0002.0013
50 4	.1006.0040.0150	.3450.0206.0670	.1297.0079 -	.0732.0064.0143	.2309.0331.0588	.1314.0053.0207
100 4	.0449.0020.0068	.1608.0099.0318	.0567.0037 -	.0358.0029.0063	.1117.0163.0278	.0600.0026.0096
200 4	.0212.0010.0032	.0726.0047.0145	.0274.0018 -	.0176.0014.0030	.0536.0075.0127	.0269.0012.0045
500 4	.0085.0004.0012	.0282.0019.0056	.0104.0007 -	.0070.0005.0012	.0210.0028.0049	.0105.0005.0017
1000 4	.0042.0002.0006	.0137.0009.0027	.0053.0004 -	.0035.0003.0006	.0104.0014.0024	.0052.0002.0008
50 5	.0745.0039.0109	.2577.0192.0522	.0908.0072 -	.0665.0047.0109	.2065.0269.0506	.0908.0050.0148
100 5	.0363.0020.0052	.1180.0096.0241	.0442.0034 -	.0333.0023.0051	.1006.0126.0232	.0431.0025.0069
200 5	.0179.0010.0025	.0574.0046.0115	.0220.0017 -	.0166.0011.0025	.0495.0059.0110	.0211.0012.0034
500 5	.0072.0004.0010	.0231.0018.0046	.0084.0007 -	.0067.0004.0010	.0197.0023.0044	.0084.0005.0013
1000 5	.0036.0002.0005	.0112.0009.0022	.0043.0003 -	.0034.0002.0005	.0098.0011.0022	.0042.0002.0006

When SMD is 1, the MSE of  $\psi_{11}$  in model A.8 clearly decreases when the sample size increases from  $n = 50$  to  $n = 1000$  (see Figure 6.23). For  $\psi_{11}$ , the MSE

decreases by one third when the sample size grows from about 50 to 95, from 100 to 190, from 200 to 420, or from 500 to 890. When SMD is 2, the MSE of  $\psi_{11}$  decreases by one third when the sample size grows from about 50 to 90, from 100 to 170, from 200 to 330 or from 500 to 850. As can be seen from Figure 6.23, the MSE of  $\psi_{11}$  is almost the same with all sample sizes when SMD is 3, 4 or 5, decreasing by half when the sample size grows from about 50 to 100, from 100 to 190-200, from 200 to 440 or from 500 to 1000.

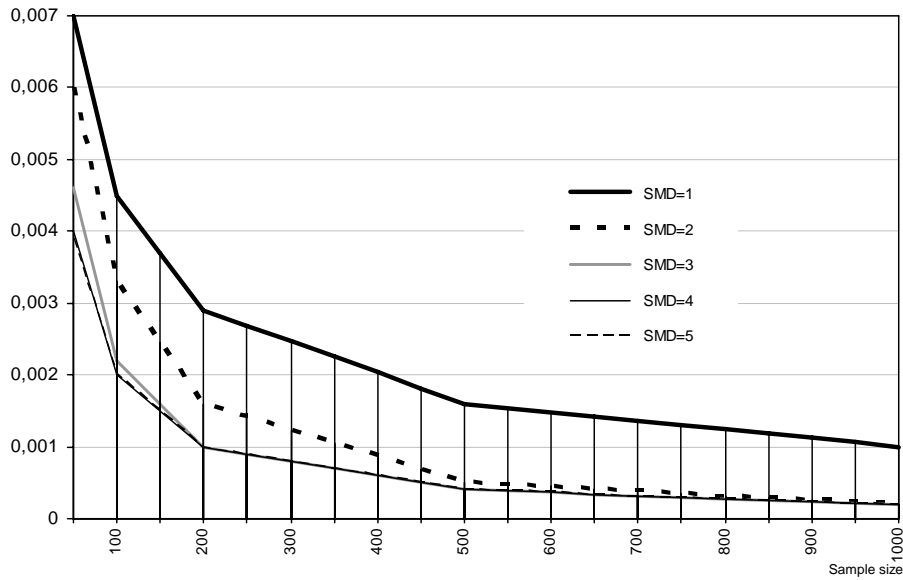


Figure 6.23. The MSE of the  $\psi_{11}$  parameter estimation as a function of sample size, SMD=1, 2, 3, 4 and 5.

When SMD is 1, the MSE of  $\psi_{01}$  in model A.8 clearly decreases when the sample size increases from  $n = 50$  to  $n = 1000$ . For  $\psi_{01}$  parameter, the MSE decreases by one third when the sample size grows from about 50 to 120, from 100 to 230, from 200 to 450, or from 500 to over 1000 (see Figure 6.24). When SMD is 2, the MSE of  $\psi_{01}$  decreases by one third when the sample size grows from about 50 to 100, from 100 to 180, from 200 to 380 or from 500 to over than 1000. When SMD is 3, the MSE of  $\psi_{01}$  decreases by half when the sample size grows from about 50 to 105, from 100 to 190, from 200 to 380 or from 500 to 980. When SMD is 4, the MSE of  $\psi_{01}$  decreases by half when the sample size grows from about 50 to 95, from 100 to 195, from 200 to 430 or from 500 to 1000. When SMD is 5, the MSE of  $\psi_{01}$  parameter decreases by half when the sample size grows from about 50 to 105, from 100 to 190, from 200 to 450 or from 500 to 1000.

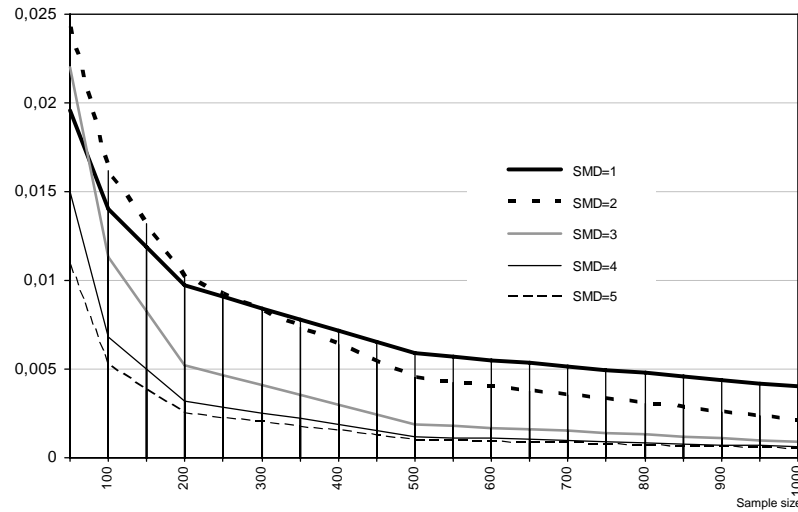


Figure 6.24. The MSE of the  $\psi_{01}$  parameter estimation as a function of sample size, SMD=1, 2, 3, 4 and 5.

The effect of SMD, when comparing the MSE of  $\psi_{00}$  with SMD 1, 2, 3 or 4 to the MSE with SMD 5 with equal sample sizes, on the MSE of  $\psi_{00}$  in model A.8, is strong. When  $n = 50$ , the MSE of  $\psi_{00}$  is 3.19, 2.24, 1.98 or 1.35 times larger when SMD is 1, 2, 3 or 4, respectively, than when SMD is 5. When SMD is 1, this proportional difference increases monotonically to 8.08 –fold when the sample size increases to  $n = 1000$ . When SMD is 2, this proportional difference increases to 4.43 –fold when the sample size increases to  $n = 200$  and then decreases to 4.06 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportional difference first increases to 2.12 when the sample size increases to  $n = 100$  and, after that, decreases to 1.17 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportional difference decreases to 1.17 when the sample size increases to  $n = 1000$ . The effect of SMD, comparing the MSE with SMD 1, 2, 3 or 4 to the MSE with SMD 5 with equal sample sizes, on the MSE in model A.8, is weaker for  $\psi_{11}$  than for  $\psi_{00}$ . When  $n = 50$ , the MSE of  $\psi_{11}$  is 1.79, 1.54, 1.18 or 1.03 times larger when SMD is 1, 2, 3 or 4, respectively, than when SMD is 5. When SMD is 1, this proportional difference increases monotonically to 5.0 –fold when the sample size increases to  $n = 1000$ . When SMD is 2, 3 or 4, the proportional differences decrease and the MSE of  $\psi_{11}$  is equal with the MSE of  $\psi_{11}$  when SMD is 5.

The effect of SMD, when comparing the MSE with SMD 1, 2, 3 or 4 to the MSE with SMD 5 with equal sample sizes, on the MSE of  $\psi_{01}$  in model A.8, is strong.

When  $n = 50$ , the MSE of  $\psi_{01}$  is 1.80, 2.23, 2.02, 1.38 times larger when SMD is 1, 2, 3 or 4, respectively, than when SMD is 5. When SMD is 1, this proportional difference increases monotonically to 8.0 –fold when the sample size increases to  $n = 1000$ . When SMD is 2, this proportional difference increases to 4.50 –fold when the sample size increases to  $n = 500$  and decreases to 4.20 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportional difference first increases to 2.18 when the sample size increases to  $n = 100$  and, after that, decreases to 1.80 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportional difference decreases to 1.20 when the sample size increases to  $n = 1000$ .

These results for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameter estimates means that the convergence is clearly lower when SMD is 1 than in the case when SMD is 5. When SMD is 2 or 3, the convergence of  $\psi_{00}$  and  $\psi_{01}$  parameter estimates is slower with smaller sample sizes  $n = 50, 100$  or  $200$ , but faster when the sample size increases from  $n = 500$  to  $n = 1000$ , when compared with the case when SMD is 5. The convergence of  $\psi_{00}$  and  $\psi_{01}$  parameter estimates is faster when SMD is 4 than when SMD is 5. The convergence of  $\psi_{11}$  parameter estimate is faster when SMD is 2, 3 or 4, than when SMD is 5. When the sample size is  $n = 1000$ , the MSE of  $\psi_{11}$  parameter is equal to the cases when SMD is 2, 3, 4 or 5.

*The effect of reliability on the MSE of variances  $\psi_{00}$ ,  $\psi_{11}$  and covariance  $\psi_{01}$  parameter estimation*

The effect of reliability on the MSE of  $\psi_{00}$ ,  $\psi_{11}$  or  $\psi_{01}$  is examined comparing the MSE in model A.5 with the MSE in model A.8, using proportions of MSEs in these models. As for intercept parameters, these proportions are calculated by dividing the MSE in model A.5 by the MSE in model A.8. When reliability of the observed variables decreases from .80 to .50 (model A.8 vs. model A.5), the MSEs of  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameter estimates strongly increase (see Table 6.13).

When  $n = 50$ , the MSE of  $\psi_{00}$  in model A.5 is 2.47, 2.72, 2.93, 3.43 or 3.46 times larger when SMD is 1, 2, 3, 4 or 5, respectively, than the MSE of  $\psi_{00}$  in model A.8. When SMD is 1, this proportional difference increases to 2.54 when sample size increases to  $n = 500$  and decreases to 2.47 when the sample size increases to  $n = 1000$ . When SMD is 2, the proportional difference decreases to 2.34 when the sample size increases to  $n = 200$  and increases then to 3.51 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportional difference increases to 3.62 when the sample size increases to  $n = 200$  and decreases then to 3.33 when

the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportional difference decreases to 3.26 or 3.11 when the sample size increases to  $n = 1000$ .

When  $n = 50$ , the MSE of  $\psi_{11}$  in model A.5 is 4.27, 4.67, 5.17, 5.15 or 4.92 times larger when SMD is 1, 2, 3, 4 or 5, respectively, than the MSE of  $\psi_{11}$  in model A.8. When SMD is 1, this proportional difference decreases to 4.0 when the sample size increases to  $n = 1000$ . When SMD is 2, the proportional difference increases to 6.5 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportional difference decreases to 4.75 when the sample size increases to  $n = 500$  and increases to 5.0 when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportional difference decreases to same value 4.50 when the sample size increases to  $n = 1000$ .

When  $n = 50$ , the MSE of  $\psi_{01}$  in model A.5 is 3.66, 3.46, 3.81, 4.47 or 4.79 times larger when SMD is 1, 2, 3, 4 or 5, respectively, than the MSE of  $\psi_{01}$  in model A.8. When SMD is 1, this proportional difference decreases to 2.95 when the sample size increases to  $n = 1000$ . When SMD is 2, the proportional difference decreases to 3.25 when the sample size increases to  $n = 200$  and increases then to 4.0 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportional difference increases to 4.35 when the sample size increases to  $n = 200$  and decreases then to 4.33 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportional difference increases to 4.68 when the sample size increases to  $n = 100$  and decreases then to 4.50 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportional difference decreases to 4.40 when the sample size increases to  $n = 1000$ .

The results means that the effect of reliability on the MSE of  $\psi_{00}$ ,  $\psi_{11}$  or  $\psi_{01}$  parameter estimates are strong. When SMD is 5 and  $n = 1000$ , the MSE of  $\psi_{00}$ ,  $\psi_{11}$  or  $\psi_{01}$  parameter estimates in model A.5 are 3.1, 4.5 or 4.4 times larger than in model A.8. The difference between the convergence rates behaves regularly only when SMD is 4 or 5, and for  $\psi_{00}$ ,  $\psi_{11}$  or  $\psi_{01}$  parameter estimates, the convergence is slightly more rapid in model A.8 than in model A.5. A small difference in the convergence rates is seen with all SMD=1, 2, 3, 4 or 5, and the largest difference in the MSE of  $\psi_{00}$ ,  $\psi_{11}$  or  $\psi_{01}$  parameter is 3.6, 6.5 or 4.8, respectively.

*The effect of additional measurements on the MSE of variances  $\psi_{00}$  and  $\psi_{11}$  parameter estimation*

The effect of additional measurement points on the MSE of  $\psi_{00}$  and  $\psi_{11}$  is examined by comparing the MSE in model A.5\* to the MSE in model A.5. As before for the intercept and slope parameters, the effect of additional measurement points is described as percentages of distance (see Table 6.13).

The MSE of  $\psi_{00}$  parameter is clearly lower in model A.5\* than in model A.5. When SMD is 1, the distance decreases from 85 % to 70 % when the sample size increases from  $n = 50$  to  $n = 1000$ . When SMD is 2, the distance increases from 73 % to 80 % when the sample size increases from  $n = 50$  to  $n = 1000$ . When SMD is 3, the distance increases from 74 % to 86 % when the sample size increases from  $n = 50$  to  $n = 500$ , and decreases after that to 84 % when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the distance varies slightly and is 88 % or 91 %, respectively, when the sample size is  $n = 1000$ .

The MSE of  $\psi_{11}$  parameter is clearly lower in model A.5\* than in model A.5. When SMD is 1, the distance increases from 57 % to 90 % when the sample size increases from  $n = 50$  to  $n = 1000$ . When SMD is 2, the distance increases from 65 % to 85 % when the sample size increases from  $n = 50$  to  $n = 500$ , and decreases after that to 82 % when the sample size increases to  $n = 1000$ . When SMD is 3, the distance increases from 71 % to 75 % when the sample size increases from  $n = 50$  to  $n = 1000$ . When SMD is 4, the distance increases from 77 % to 80 % when the sample size increases from  $n = 50$  to  $n = 500$ , and decreases after that to 71 % when the sample size increases to  $n = 1000$ . When SMD is 5, the distance varies and increases from .78 % to .86 % when the sample size increases from  $n = 50$  to  $n = 1000$ .

The convergence rate of  $\psi_{00}$  and  $\psi_{11}$  parameters is very similar in model A.5\* than in model A.5, in spite of that the MSE in model A.5\* is near to the MSE in model A.8 and the difference from the MSE in model A.5 is at least 57 %. The convergence rate of  $\psi_{00}$  parameter is slightly more rapid in model A.5\* than in model A.5 when SMD was 3. The convergence rate of  $\psi_{11}$  parameter is slightly more rapid in model A.5\* than in model A.5 when SMD was 2.

*The effect of model construct on the MSE of variances  $\psi_{00}$ ,  $\psi_{11}$  and covariance  $\psi_{01}$  parameter estimation*

The estimation of parameter  $\psi_{00}$  is more effective in model A.8 than in model C.8 (see Table 6.13), and when  $n = 50$ , the MSE in model C.8 is 1.10, 1.08, 1.18, 1.31 and 1.22 times larger than the MSE in model A.8 when SMD is 1, 2, 3, 4 or 5, respectively. When SMD is 1, the proportion increases to 1.14 when the sample size increases to  $n = 500$ , and decreases after that to 1.11 when the sample size increases to  $n = 1000$ . When SMD is 2, the proportion decreases to 1.02 when the sample size increases to  $n = 200$  and increases after that to 1.57 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion increases to 1.45 when the sample size increases to  $n = 200$ , and decreases after that to 1.35 when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportion decreases to 1.24 or 1.17, respectively, when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\psi_{00}$  is more effective in model B.8 than in model A.8 (see Table 6.13), and when  $n = 50$ , the MSE in model B.8 is .81, .64, .73 or .89 times the MSE in model A.8 when SMD is 2,3,4 or 5, respectively. When SMD is 2, the proportion decreases to .62 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion increases to .70 when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportion increases to .83 or .94, respectively, when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\psi_{00}$  is more effective in model B.5 than in model A.5 (see Table 6.13), and when  $n = 50$ , the MSE in model B.5 is .78, .66, .67 or .80 times the MSE in model A.5 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, the proportion decreases to .52 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion decreases to .58 when the sample size increases to  $n = 200$  and, after that, increases to .63 when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportion increases to .76 or .88, respectively, when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\psi_{11}$  is more effective in model A.8 than in model C.8, and when  $n = 50$ , the MSE in model C.8 is 1.20, 1.32, 1.37, 1.33 and 1.28 times larger than the MSE in model A.8 when SMD is 1, 2, 3, 4 or 5, respectively. When SMD is 1 or 2, the proportion increases to 1.30 or 2.0, respectively, when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion increases to 1.45 when the sample size increases to  $n = 200$  and, after that, decreases to 1.415 when the sample size increases to  $n = 500$  and then decreases again to 1 when the



sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportion decreases to 1.0 when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\psi_{11}$  is more effective in model A.8 than in model B.8, and when  $n = 50$ , the MSE in model B.8 is 1.61, 1.98, 1.60 or 1.21 times larger than the MSE in model A.8 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, the proportion increases to 4.0 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion increases to 2.20 when the sample size increases to  $n = 200$  and, after that, decreases to 2.0 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportion decreases to 1.25 when the sample size increases to  $n = 500$  and, after that, increases to 1.50 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion decreases to 1.0 when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\psi_{11}$  is more effective in model A.5 than in model B.5, and when  $n = 50$ , the MSE in model B.5 is 1.35, 1.59, 1.61 or 1.40 times larger than the MSE in model A.5 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, the proportion increases to 2.69 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion increases to 2.22 when the sample size increases to  $n = 200$  and, after that, decreases to 2.0 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportion decreases to 1.47 when the sample size increases to  $n = 500$  and, after that, increases to 1.56 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion decreases to 1.22 when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\psi_{01}$  is more effective in model A.8 than in model C.8, and when  $n = 50$ , the MSE in model C.8 is 1.11, 1.10, 1.21, 1.38 and 1.36 times larger than the MSE in model A.8 when SMD is 1, 2, 3, 4 or 5, respectively. When SMD is 1, the proportion is very stable and is 1.10 when  $n = 1000$ . When SMD is 2, the proportion increases to 1.57 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion increases to 1.48 when the sample size increases to  $n = 200$  and, after that, decreases to 1.44 when the sample size increases to  $n = 1000$ . When SMD is 4, the proportion increases to 1.42 when the sample size increases to  $n = 500$  and, after that, decreases to 1.33 when the sample size increases to  $n = 1000$ . When SMD is 5, the proportion decreases to 1.20 when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\psi_{01}$  is more effective in model B.8 than in model A.8, and when  $n = 50$ , the MSE in model B.8 is .74, .84, .95 or 1.0 times the MSE in model A.8 when SMD is 2,3,4 or 5, respectively. When SMD is 2, the proportion

increases to .75 when the sample size increases to  $n = 100$  and, after that, decreases to .52 when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion decreases to .78 when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportion varies slightly and is 1.0 when  $n = 1000$ .

The estimation of parameter  $\psi_{01}$  is more effective in model B.5 than in model A.5, and when  $n = 50$ , the MSE in model B.5 is .75, .78, .88 or .97 times the MSE in model A.5 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2 or 3, the proportion decreases to .52 or .69, respectively, when the sample size increases to  $n = 1000$ . When SMD is 4, the proportion is very stable and is .89 when  $n = 1000$ . When SMD is 5, the proportion increases to 1.0 when the sample size increases to  $n = 1000$ .

#### 6.3.1.4. Results of MSE of error variances $\theta_1, \theta_2, \theta_3$ and $\theta_4$ estimation

The results of the MSE for  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameter estimation are presented in Table 6.14. The running order is the same as for other parameters in sections 6.3.1.1 - 6.3.1.3. The results suggest that in model A.8 the value of the MSE depends only on sample size. The effect of sample size on  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameter estimates in model A.8 when SMD is 3 is described in figure 6.25.

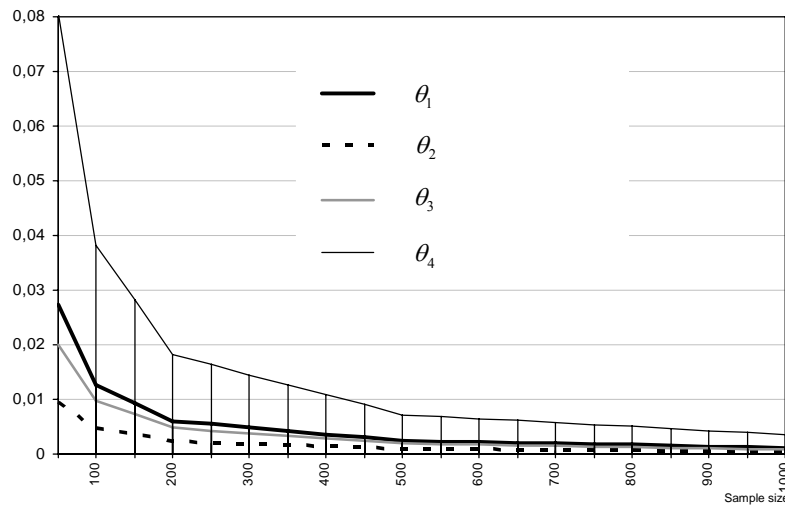


Figure 6.25. The MSE of  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameter estimation as a function of sample size in model A.8, SMD=3.

When SMD is 3, the MSE of  $\theta_1, \theta_2, \theta_3$  or  $\theta_4$  in model A.8 decreases clearly when the sample size increases from  $n = 50$  to  $n = 1000$ . The MSE of these parameter estimates decreases by half when the sample size grows from about 50 to 95-100, from 100 to 195-200, from 200 to 440-450, or from 500 to 940-1000, respectively.

*The effect of reliability on the MSE of error variances  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  estimation*

The effect of reliability on the MSE of  $\theta_1, \theta_2, \theta_3$  or  $\theta_4$  is examined by comparing the MSEs in model A.5 to the MSEs in model A.8, using the proportions of the MSE. As before for other parameters, these proportions are calculated by dividing the MSE of model A.5 by the MSE of model A.8 (see Table 6.14). For error variances, the proportions of MSEs are not absolutely comparable, because of different true values. For the first measurement, the proportions are between 8.5 and 9.3, for the second measurement, they are between 12.2 and 13.5, for the third, between 12.2 and 12.7, and for the last measurement, between 10.6 and 11.0. For error variances, the proportions are very stable suggesting that the convergence is nearly the same in model A.5 as in model A.8.

*The effect of additional measurements on the MSE of error variances  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  estimation*

The effect of additional measurement points on the MSE of  $\theta_1, \theta_2, \theta_3$  or  $\theta_4$  is examined by comparing the MSEs in model A.5\* with the MSEs in model A.5 (see Table 6.14). As before for other parameters, the effect of additional measurement points is described as percentages of distance. The MSEs of  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameters in model A.5\* are clearly closer with the MSEs in model A.5 than the MSEs in model A.8. The distances from the MSEs of these parameters in model A.5 are very stable, and when SDM is 5 and  $n = 1000$ , they are 64 %, 25 %, 22 % and 38 %, respectively.

*The effect of model construct on the MSE of error variances  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  estimation*

The estimation of parameter  $\theta_1$  is more effective in model A.8 than in model C.8 (see Table 6.14), and when  $n = 50$ , the MSE in model C.8 is 1.02, 1.08, 1.16, 1.19 and 1.14 times larger than the MSE in model A.8 when SMD is 1, 2, 3, 4 or 5, respectively. When SMD is 1, the proportion is stable decreasing to 1.0 when the sample size increases to  $n = 1000$ . When SMD is 2, the proportion decreases to 1.05 when the sample size increases to  $n = 100$  and, after that, increases to 1.09

Table 6.14. The MSEs of  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameter estimation in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8				A.5				A.5*1)				B.8				B.5				C.8			
n	SMD	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_3$	$\theta_5$	$\theta_7$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
50	1	.0251	.0091	.0203	.0831	.2160	.1139	.2508	.9002	.0978	.0836	.1935	.5525	.0251	.0091	.0203	.0831	.2160	.1139	.2508	.9002	.0257	.0149	.0405	.1521
100	1	.0113	.0043	.0098	.0391	.0978	.0550	.1211	.4154	.0453	.0406	.0456	.2640	.0113	.0043	.0098	.0391	.0978	.0550	.1211	.4154	.0114	.0072	.0194	.0707
200	1	.0054	.0021	.0048	.0184	.0469	.0264	.0586	.1946	.0221	.0201	.0456	.1308	.0054	.0021	.0048	.0184	.0469	.0264	.0586	.1946	.0054	.0035	.0095	.0333
500	1	.0021	.0008	.0019	.0073	.0178	.0103	.0233	.0779	.0086	.0080	.0181	.0523	.0021	.0008	.0019	.0073	.0178	.0103	.0233	.0779	.0021	.0014	.0038	.0132
1000	1	.0010	.0004	.0009	.0036	.0089	.0051	.0114	.0391	.0044	.0040	.0091	.0258	.0010	.0004	.0009	.0036	.0089	.0051	.0114	.0391	.0010	.0007	.0018	.0067
50	2	.0278	.0097	.0201	.0823	.2396	.1220	.2530	.8981	1.006	.0847	.1933	.5486	.0217	.0085	.0226	.0977	.1921	.1109	.2852	1.0703	.0301	.0158	.0405	.1528
100	2	.0126	.0047	.0098	.0388	.1085	.0589	.1222	.4161	.0470	.0412	.0950	.2650	.0102	.0041	.0109	.0451	.0899	.0543	.1354	.4812	.0132	.0076	.0194	.0710
200	2	.0060	.0022	.0048	.0183	.0520	.0284	.0594	.1956	.0230	.0204	.0457	.1307	.0049	.0020	.0052	.0210	.0436	.0263	.0649	.2243	.0063	.0037	.0095	.0337
500	2	.0023	.0009	.0019	.0072	.0197	.0110	.0236	.0785	.0089	.0081	.0182	.0521	.0019	.0008	.0021	.0082	.0170	.0103	.0256	.0885	.0024	.0014	.0038	.0134
1000	2	.0011	.0004	.0009	.0036	.0098	.0055	.0115	.0394	.0045	.0040	.0091	.0257	.0010	.0004	.0010	.0041	.0085	.0051	.0125	.0445	.0012	.0007	.0018	.0067
50	3	.0274	.0096	.0199	.0806	.2480	.1270	.2520	.8884	.0987	.0845	.1917	.5409	.0218	.0083	.0237	.1030	.1879	.1094	.3044	1.1523	.0319	.0162	.0398	.1495
100	3	.0127	.0047	.0097	.0382	.1138	.0612	.1222	.4156	.0461	.0414	.0949	.2626	.0103	.0041	.0113	.0480	.0892	.0538	.1450	.5242	.0143	.0078	.0193	.0708
200	3	.0061	.0023	.0048	.0182	.0546	.0294	.0594	.1958	.0225	.0250	.0457	.1302	.0050	.0020	.0054	.0223	.0432	.0260	.0686	.2447	.0068	.0038	.0095	.0336
500	3	.0024	.0009	.0019	.0072	.0207	.0114	.0236	.0788	.0088	.0082	.0182	.0520	.0020	.0008	.0021	.0087	.0171	.0102	.0269	.0958	.0026	.0015	.0038	.0134
1000	3	.0012	.0004	.0009	.0036	.0103	.0056	.0115	.0395	.0045	.0041	.0091	.0257	.0010	.0004	.0010	.0044	.0085	.0051	.0132	.0485	.0013	.0007	.0018	.0068
50	4	.0248	.0090	.0196	.0791	.2310	.1212	.2468	.8658	.0911	.0837	.1910	.5329	.0216	.0082	.0228	.0976	.1845	.1076	.3043	1.1512	.0294	.0157	.0393	.1463
100	4	.0117	.0044	.0096	.0380	.1072	.0588	.1209	.4103	.0430	.0411	.0948	.2610	.0103	.0040	.0109	.0456	.0887	.0531	.1441	.5239	.0134	.0076	.0192	.0698
200	4	.0056	.0022	.0047	.0181	.0521	.0285	.0591	.1948	.0213	.0203	.0457	.1296	.0050	.0020	.0053	.0214	.0432	.0257	.0685	.2455	.0064	.0037	.0095	.0333
500	4	.0022	.0009	.0019	.0072	.0202	.0111	.0235	.0785	.0083	.0081	.0182	.0518	.0020	.0008	.0021	.0084	.0172	.0101	.0268	.0964	.0025	.0015	.0038	.0134
1000	4	.0011	.0004	.0009	.0036	.0099	.0055	.0115	.0393	.0043	.0040	.0091	.0256	.0010	.0004	.0010	.0042	.0086	.0050	.0132	.0490	.0012	.0007	.0018	.0067
50	5	.0224	.0085	.0195	.0785	.2086	.1126	.2433	.8486	.0854	.0826	.1903	.5300	.0212	.0082	.0213	.0891	.1834	.1055	.2878	1.0660	.0256	.0147	.0389	.1432
100	5	.0107	.0042	.0096	.0380	.0982	.0554	.1195	.4050	.0406	.0406	.0947	.2604	.0102	.0040	.0102	.0422	.0882	.0524	.1364	.4912	.0119	.0072	.0191	.0690
200	5	.0052	.0021	.0047	.0181	.0481	.0270	.0586	.1931	.0201	.0201	.0456	.1292	.0049	.0020	.0050	.0199	.0432	.0255	.0659	.2315	.0057	.0035	.0094	.0330
500	5	.0021	.0008	.0019	.0072	.0190	.0106	.0233	.0779	.0079	.0080	.0181	.0518	.0020	.0008	.0020	.0078	.0172	.0101	.0259	.0917	.0023	.0014	.0038	.0133
1000	5	.0010	.0004	.0009	.0036	.0093	.0052	.0114	.0390	.0040	.0040	.0091	.0256	.0010	.0004	.0010	.0040	.0086	.0050	.0127	.0466	.0011	.0007	.0018	.0067

1)  $\theta_3 = \theta_2, \theta_5 = \theta_3, \theta_7 = \theta_4$

when the sample size increases to  $n = 1000$ . When SMD is 3, the proportion decreases to 1.08 when the sample size increases to  $n = 1000$ . When SMD is 4 or 5, the proportion decreases to 1.09 or 1.10, respectively, when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\theta_1$  is more effective in model B.8 than in model A.8, and when  $n = 50$ , the MSE in model B.8 is .78, .80, .87 or .95 times the MSE in model A.8 when SMD is 2,3,4 or 5, respectively. When SMD is 2, 3, 4 or 5, the proportion increases to .91, .83, .91 or 1.0, respectively when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\theta_1$  is more effective in model B.5 than in model A.5, and when  $n = 50$ , the MSE in model B.5 is .80, .76, .80 or .88 times the MSE in model A.5 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, 3, 4 or 5, the proportion increases to .87, .83, .87 or .92, respectively when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\theta_2$  is more effective in model A.8 than in model C.8, and when  $n = 50$ , the MSE in model C.8 is 1.64, 1.63, 1.69, 1.74 and 1.73 times larger than the MSE in model A.8 when SMD is 1, 2, 3, 4 or 5, respectively. When SMD is 1, 2, 3, 4 or 5, the proportion varies slightly, but increases in all cases to 1.75 when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\theta_2$  is more effective in model B.8 than in model A.8, and when  $n = 50$ , the MSE in model B.8 is .88, .86, .91 or .99 times the MSE in model A.8 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, 3, 4 or 5, the proportion varies slightly, but increases in all cases to 1.0 when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\theta_2$  is more effective in model B.5 than in model A.5, and when  $n = 50$ , the MSE in model B.5 is .91, .86, .89 or .94 times the MSE in model A.5 when SMD is 2,3,4 or 5, respectively. When SMD is 2, 3, 4 or 5, the proportion varies slightly, but increases to .93, .91, .91 or .96, respectively, when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\theta_3$  is more effective in model A.8 than in model C.8, and the proportion varies slightly being 1.98 – 2.02 with all sample sizes and SMDs.

The estimation of parameter  $\theta_3$  is more effective in model A.8 than in model B.8, and when  $n = 50$ , the MSE in model B.8 is 1.12, 1.19, 1.16 or 1.09 times larger than the MSE in model A.8 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, 3, 4 or 5, the proportion varies slightly, but decreases in all cases to 1.11 when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\theta_3$  is more effective in model B.5 than in model A.5, and when  $n = 50$ , the MSE in model B.5 is 1.13, 1.21, 1.23, 1.18 times larger than the MSE in model A.5 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, 3, 4 or 5, the proportion varies slightly and decreases to 1.09, 1.15, 1.15 or 1.11, respectively, when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\theta_4$  is more effective in model A.8 than in model C.8, and the proportion varies slightly being 1.81 – 1.86 with all sample sizes and SMDs.

The estimation of parameter  $\theta_4$  is more effective in model A.8 than in model B.8, and when  $n = 50$ , the MSE in model B.8 is 1.19, 1.28, 1.23 or 1.14 times larger than the MSE in model A.8 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, 3, 4 or 5, the proportion varies slightly, but decreases to 1.14, 1.22, 1.17 or 1.11, respectively, when the sample size increases to  $n = 1000$ .

The estimation of parameter  $\theta_4$  is more effective in model B.5 than in model A.5, and when  $n = 50$ , the MSE in model B.5 is 1.19, 1.30, 1.33, 1.26 times larger than the MSE in model A.5 when SMD is 2, 3, 4 or 5, respectively. When SMD is 2, 3, 4 or 5, the proportion varies slightly but decreases to 1.13, 1.23, 1.25 or 1.19, respectively, when the sample size increases to  $n = 1000$ .

### **6.3.1.5. Summary of the results of MSE**

#### *The effects of sample size and SMD*

The MSE decreases for all parameters and in all models as a function of sample size when SMD is large enough. The effect of SMD on the MSE seems to be very strong. The comparison of MSEs when SMD is large;  $SMD = 4$  or  $5$ , reveals that the effect of SMD disappears when SMD increases.

When SMD is 1, the MSEs of  $\alpha_0^{(1)}$ ,  $\alpha_0^{(2)}$ ,  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter estimates in model A.8 slowly decrease, when the sample size increases. When the sample size

is 50, the MSE of these parameter estimates is 5.0-11.8 times larger when SMD is 1, than when compared to the situation where SMD is 5. The difference between the MSEs in these two situations increases when the sample size increases. When  $n = 1000$ , the MSE is 69.1 - 88.9 larger when SMD is 1, than when compared to the situation where SMD is 5. In turn, the MSEs of  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameter estimates show a clear decrease when the sample size increases. When the sample size is 50, the MSEs of these parameter estimates are 1.79-3.19 times larger when SMD is 1, than when compared to the situation where SMD is 5. This difference between the MSEs increases being 5.0-8.08 when the sample size increases to 1000. The MSEs of  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ ,  $\theta_4$  parameter estimates are almost equal to all SMDs and the respective sample sizes.

When SMD is 2, the MSE of all parameter estimates in model A.8 decreases by one third when the sample size increases by 1.6 – 2.4 times. When the sample size is small ( $n = 50$  or  $100$ ), the MSE decreases most rapidly for  $\alpha_0^{(1)}$ ,  $\alpha_0^{(2)}$  and  $\psi_{11}$  parameters in which case the MSE decreases by one third when the sample size increases by 1.6 – 1.8 times. For  $\alpha_1^{(2)}$  parameter, in turn, the sample size must increase by 2.0 -2.4 times in order to achieve the same decrease in the MSE. When  $n \geq 200$ , the decrease in the MSE is rapid for  $\alpha_0^{(2)}$ ,  $\alpha_1^{(1)}$  and  $\psi_{11}$  parameters. This decrease is one third when the sample size increases by 1.5 - 1.7 times. When  $n \geq 500$ , the decrease in the MSE is also rapid for  $\alpha_0^{(1)}$ ,  $\alpha_1^{(2)}$  and  $\psi_{00}$  parameters. For these parameters, the MSE decreases by one third when the sample size increases by 1.5 - 1.6 times. The decrease in the MSE is slowest for  $\psi_{01}$  parameter. When  $n = 50, 100$ , or  $200$ , the MSE of  $\psi_{01}$  decreases by one third when the sample size increases by 1.8 – 1.9 times, and when  $n = 500$ , the sample size must increase by over two times.

When SMD is 3, 4 or 5, the MSE of all parameter estimates in model A.8 decreases by half when the sample size increases by two times. This amount of decrease is seen for all mean parameters, for the variance of intercept, for the covariance between intercept and slope, and for the variance of slope. This amount of decrease is seen for all error variances ( $\theta_1, \theta_2, \theta_3, \theta_4$ ) also in the case when SMD is 1 or 2.

The pattern of convergence is similar with different values of SMD=2, 3, 4 or 5, but the effect of SMD on the MSE is large for all other parameters except for  $\psi_{11}$ ,  $\theta_1, \theta_2, \theta_3, \theta_4$ . When  $n = 1000$ , the MSE is 4 – 6 times larger for  $\alpha_0^{(1)}$ ,  $\alpha_0^{(2)}$ ,  $\alpha_1^{(1)}$ ,  $\alpha_1^{(2)}$ ,  $\psi_{00}$ ,  $\psi_{01}$  parameters when SMD is 2, when compared with the case when

SMD is 5. If  $n = 50$ , the MSE of these parameters is 1.7- 2.5 times larger when SMD is 3 than when SMD is 5. If  $n = 1000$ , the MSEs of  $\alpha_0^{(1)}$ ,  $\alpha_0^{(2)}$ ,  $\alpha_1^{(2)}$  and  $\psi_{01}$  parameter estimates are 1.6 – 1.8 larger, and those of other parameters 1.0 - 1.25 larger, in the situation where SMD is 3, when compared with the situation where SMD is 5. The MSE of  $\alpha_0^{(2)}$  or  $\alpha_1^{(2)}$  is about two times larger than the MSE of  $\alpha_0^{(1)}$  or  $\alpha_1^{(1)}$ , respectively. This result is obvious, because the expected size of class 1 is two times larger than that of class 2.

### *The effect of reliability*

When reliability of observed variables decreases from .80 to .50 (model A.8 vs. model A.5), the MSEs of all parameter estimation increase. The pattern of the decreasing MSE with different SMDs is similar in model A.5 as in model A.8. Larger proportions between the MSEs when comparing model A.5 with model A.8 are most obvious when SMD is 2 or 3. For example, when SMD is 3 and  $n = 200$ , the MSEs are 2.1 – 3.0 times larger for mean parameters, and 3.6 – 5.1 times larger for the variances and covariance of latent components, in model A.5 than in model A.8. When SMD is 5, the MSEs are 1.6 – 2.0 times larger for mean parameters, and 3 - 4.9 times larger for the variances and covariance of latent components, in model A.5 than in model A.8. These differences between the MSEs are very stable with different sample sizes. These results mean that when the sample size increases to two or five times larger in model A.5, the MSEs of mean parameters, or the MSEs of variances and covariance of latent components, decrease to the same level as these values in model A.8.

### *The effect of additional measurement points*

The effect of additional measurement points (described as percentage of distance, presented in equation 6.1) on the MSE is, on average, strong and usually greater than 50%. However, this effect is weak for the mean of latent components when SMD is 1 or when SMD is 2 and the sample size is small. When SMD is 2 and the sample size is 1000, the effect on  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  is 58% or 79%, respectively, and on  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  109% or 105%, respectively. This result suggests that, when the sample size is large and SMD is 2, additional measurement points have a very strong effect, especially, on the MSE of mean of slope. This kind of effect on mean parameters is shown also when SMD is 3, although this effect is in this case slightly weaker. When SMD is 4 or 5, this effect becomes weaker and is on  $\alpha_0^{(1)}$ ,  $\alpha_0^{(2)}$ ,  $\alpha_1^{(1)}$ ,  $\alpha_1^{(2)}$  only 36%, 43%, 25% or 44%, respectively, when the sample size is 1000.



The effect of additional measurement points on the MSE of  $\psi_{00}$  is strong and rather stable. When SMD increases, this effect slightly increases and is between 73 – 91 %. The effect of additional measurement points on the MSE of  $\psi_{11}$  increases from 57% to 90%, or from 65% to 82%, when SMD is 1 or 2, respectively, and is rather stable (71-86%) when SMD is 3, 4 or 5. The effect on the MSE of error variances is very stable. This effect is strongest on the error variance of the first measurement, 64%, and on the error variance of the last measurement, 38%, whereas the effect is weak on error variances of the second and third measurements, 25% and 22%, respectively.

### *The effect of construct*

The construct of the model has a clear effect on the MSE. This was examined by comparing model C.8 with model A.8, model B.8 with model A.8, and model B.5 with A.5, respectively. The results suggest, first, that the MSEs in model C.8 are greater than the MSEs in model A.8 for all parameters. The MSEs of parameters in model C.8 are between the MSEs in model A.8 and A.5. When SMD is large and the sample size increases, the MSEs in model C.8 approaches the MSEs in model A.8. The difference in the MSE is clearest, on average, when SMD is 3 and the sample size is 200. In this case, the MSEs are 1.5 - 1.8 times larger for mean parameters, and 1.4 – 1.5 times larger for the variance and covariance parameters of latent components, in model C.8 than in model A.8. When SMD is 5 and the sample size is 1000, the MSEs of latent component parameters are about 1.1 – 1.2 times larger in model C.8 than in model A.8, except that the MSE of variance of slope in model C.8 is equal to the respective MSE in model A.8.

Second, the MSE in model B.8 or B.5 is smaller than the MSE in model A.8 or A.5, respectively, for the mean of intercept component, the variance of intercept component, the covariance of intercept and slope components, and the first and second error variances. For other parameters, the MSE is larger in model B.8 and B.5 than in model A.8 and A.5, respectively. The differences in the MSEs are greater between models B.5 and A.5 than between models B.8 and A.8, with few exceptions. The differences in the MSEs are most evident when SMD is 3. When the sample size is 1000 and SMD is 3 or 5, the MSE is about 13-22 or 3-7 percent lower for the mean of intercept, 34-46 or 11-28 percent greater for the mean of slope, 30-37 or 12-20 percent lower for the variance of intercept, 100 or 0-22 percent greater for the variance of slope, and 22-31 or near to zero for the covariance of intercept and slope, respectively, in model B.5 than in model A.5. The MSE of the first and second error variance is 4 -8 percent lower, and the MSE of the third and fourth error variances 11 - 19 percent larger, in model B.5 than in model A.5.

### 6.3.2. Results of proportion of parameter bias in MSE (PB)

The next chapter describes what is the proportion of parameter bias in MSE (PB). As for MSE, the results of PB are first expressed in A.8 model for sample size  $n$  and SMD. If the PB is lower than .01, the proportion of parameter bias is one percent and is seen to be negligible when estimating the true value of parameter.

#### 6.3.2.1. Results of PB for $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$

##### *The effects of sample size and SMD*

As can be seen from Table 6.15., the PB in model A.8 is large when SMD is 1 and for  $\alpha_0^{(1)}$ , the PB decreases from .403 percent to .105 percent when the sample size increases from  $n=50$  to  $n=1000$ . The cells of Table 6.15 are highlighted in grey if the PB is greater than .01. When SMD is 2, the PB decreases from .096 to .037 when the sample size increases from  $n=50$  to  $n=100$  and, when  $n \geq 200$ , the PB is lower than .01. When  $SMD \geq 3$ , the PB is lower than .01 with all sample sizes  $n=50, 100, 200, 500$  or  $1000$ . When SMD is 1, the PB for  $\alpha_0^{(2)}$  decreases from .170 to .054 when the sample size increases from 50 to 1000. When SMD is 2, the PB is .010 when  $n=50$ , and is lower than .01 when  $n \geq 100$ . When SMD is 3, 4 or 5, the PB for  $\alpha_0^{(2)}$  is lower than .01 with all sample sizes.

##### *The effect of reliability on the PB for $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$*

The effect of reliability on the PB for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  is small and the PB in model A.5 is slightly greater than in model A.8. As can be seen from Table 6.15, the PB in model A.5 is large when SMD is 1, and for  $\alpha_0^{(1)}$ , the PB decreases from .410 to .147 when the sample size increases from 50 to 1000 and the difference to the PB in model A.8 increases from .008 to .042. When SMD is 2, the PB in model A.5 decreases from .146 to .024 when the sample size increases from 50 to 200 and the difference to the PB in model A.8 decreases from .012 to .002, respectively. As in model A.8, in model A.5, the PB is lower than .01 when  $SMD \geq 3$  with all sample sizes, except when SMD is 3 and  $n=50$ , the PB in model A.5 is .024.

For  $\alpha_0^{(2)}$ , when SMD is 1, the PB decreases in model A.5 from .204 to .066 when the sample size increases from 50 to 1000. The difference to the PB in model A.8 decreases from .034 to .012, respectively. When SMD is 2, the PB in model A.5

decreases from .023 to lower than .013 when the sample size increases from 50 to 100 and the difference to the PB in model A.8 decreases from .013 to .006. As in model A.8, the PB is lower than .01 when  $n \geq 200$ . When  $SMD \geq 3$ , as in model A.8, in model A.5, the PB is lower than .01 when  $SMD \geq 3$  with all sample sizes.

Table 6.15. The PB for  $\alpha_0^{(1)}, \alpha_0^{(2)}$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8		A.5		A.5*		B.8		B.5		C.8	
n	SMD	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$
50	1	.4028	.1699	.4103	.2040	.3347	.1175	.4028	.1699	.4103	.2040	.4364	.2249
100	1	.3341	.1246	.3474	.1500	.2862	.0946	.3341	.1246	.3474	.1500	.3770	.1664
200	1	.2612	.0886	.2856	.1161	.2367	.0793	.2612	.0886	.2856	.1161	.3008	.1271
500	1	.1738	.0619	.2048	.0802	.1659	.0705	.1738	.0619	.2048	.0802	.2318	.0877
1000	1	.1054	.0537	.1473	.0657	.1104	.0618	.1054	.0537	.1473	.0657	.1586	.0720
50	2	.0961	.0101	.1455	.0233	.1021	.0103	.0240	.0029	.0415	.0073	.1592	.0240
100	2	.0372	.0067	.0720	.0130	.0415	.0119	.0085	.0021	.0190	.0040	.0834	.0153
200	2	.0068	.0036	.0243	.0091	.0099	.0143	.0011	.0010	.0051	.0025	.0274	.0119
500	2	.0003	.0012	.0005	.0057	.0000	.0147	.0000	.0004	.0001	.0014	.0015	.0063
1000	2	.0000	.0016	.0002	.0028	.0000	.0052	.0000	.0005	.0001	.0012	.0000	.0050
50	3	.0060	.0001	.0142	.0017	.0049	.0020	.0002	.0000	.0023	.0002	.0183	.0013
100	3	.0021	.0001	.0029	.0008	.0010	.0011	.0001	.0002	.0006	.0003	.0044	.0006
200	3	.0014	.0000	.0010	.0000	.0005	.0002	.0001	.0003	.0001	.0003	.0019	.0000
500	3	.0007	.0002	.0005	.0006	.0004	.0000	.0000	.0002	.0000	.0002	.0011	.0005
1000	3	.0004	.0000	.0003	.0002	.0005	.0000	.0000	.0000	.0000	.0001	.0006	.0001
50	4	.0004	.0000	.0011	.0000	.0002	.0002	.0000	.0000	.0000	.0000	.0016	.0000
100	4	.0008	.0001	.0011	.0001	.0002	.0001	.0002	.0002	.0000	.0001	.0012	.0001
200	4	.0007	.0000	.0008	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0010	.0000
500	4	.0002	.0000	.0003	.0001	.0000	.0000	.0000	.0001	.0000	.0001	.0004	.0001
1000	4	.0002	.0000	.0001	.0000	.0002	.0002	.0000	.0000	.0000	.0000	.0002	.0000
50	5	.0000	.0000	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0000
100	5	.0003	.0001	.0007	.0001	.0000	.0001	.0002	.0003	.0002	.0002	.0005	.0001
200	5	.0002	.0000	.0005	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0003	.0000
500	5	.0000	.0000	.0001	.0001	.0000	.0000	.0000	.0001	.0000	.0001	.0001	.0000
1000	5	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0001	.0001

Note. The cell is highlighted in grey when the proportion of bias in the MSE is greater than one percent.

*The effect of additional measurements on the PB for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

The PB for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  is slightly lower in model A.5\* than in model A.5. When SMD is 1, the PB of  $\alpha_0^{(1)}$  decreases in model A.5 from .335 to .110 when the sample size increases from 50 to 1000 and the difference to the PB in model A.5 decreases from .075 to .037 percent. When SMD is 2, the PB decreases from .102 to .042 when the sample size increases from 50 to 100 and the difference decreases from .043 to .005, respectively. As in model A.5, in model A.5\*, the PB is lower than .01 when  $n \geq 200$ . As in model A.5, in model A.5\*, the PB is lower than .01 when  $SMD \geq 3$ . Other than in model A.5, the PB is also lower than .01 when  $SMD = 3$  and  $n=50$ .

For  $\alpha_0^{(2)}$ , the PB decreases in model A.5\* from .118 to .062 percent when the sample size increases from 50 to 1000 and the difference to the PB in model A.5 decreases from .087 to .004, respectively. When SMD is 2, the PB is slightly over .01 when  $n=50, 100, 200$  or  $500$  and is lower than .01 when  $n=1000$ . These values are .013 or .001 lower when  $n=50$  or  $100$ , respectively, and .004 or .015 greater when  $n=200$  or  $500$ , respectively. As in model A.5, in model A.5\*, the PB is lower than .01 when  $SMD \geq 3$  with all sample sizes.

*The effect of model construct on the PB for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

The PB in model C.8 for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  is slightly greater than the PB in model A.8. As can be seen from Table 6.15, when SMD is 1, the PB in model C.8 for  $\alpha_0^{(1)}$  decreases from .436 to .159 when the sample size increases from 50 to 1000 and the difference to the PB in model A.8 varies between .034 and .058. When SMD is 2, the PB decreases from .159 to .083 when the sample size increases from 50 to 100 and the difference to the PB in model A.8 decreases from .063 to 0.046, respectively. When  $n=200$ , the PB is .027 and is .021 greater than the PB in model A.8. As in model A.8, the PB in model C.8 is lower than .01 when  $n \geq 500$ .

As in model A.8, in model C.8, the PB is lower than .01 when  $n \geq 200$ . As in model A.8, the PB for  $\alpha_0^{(1)}$  in model C.8 is lower than .01 when  $SMD \geq 3$ . Other than in model A.8, the PB is greater than .01 and is .018 when  $SMD = 3$  and  $n=50$ . As in model A.8, the PB for  $\alpha_0^{(1)}$  in model C.8 is lower than .01 when  $SMD \geq 4$  with all sample sizes.

When SMD is 1, the PB in model C.8 for  $\alpha_0^{(2)}$  is lower than the PB in model A.8. The PB decreases from .225 to .072 when the sample size increases from 50 to 1000 and the difference to the PB in model A.8 decreases from .055 to .018 percent. When SMD is 2, the PB is .024 when  $n=50$  and is .014 greater than in model A.8. Other than in model A.8, the PB for  $\alpha_0^{(2)}$  in model C.8 is greater than .01 when  $n=100$  or  $200$ . These values are .015 and .012, respectively. As in model A.8, the PB for  $\alpha_0^{(2)}$  in model C.8 is lower than .01 when  $SMD \geq 3$  with all sample sizes.

When SMD is 1, model B.8 is equal to the model A.8. When SMD is 2, the PB in model B.8 for  $\alpha_0^{(1)}$  is .024 and is .072 lower than the PB in model A.8. Other than in model A.8, the PB for  $\alpha_0^{(1)}$  is lower than 0.1 percent when  $n=100$  and is .029 lower than the PB in model A.8. As in model A.8, the PB in model B.8 is lower than .01 when SMD is 2 and  $n \geq 200$  or when  $SMD \geq 3$  with all sample sizes.

As in model A.8, the PB for  $\alpha_0^{(2)}$  in model B.8 is lower than .01 when  $SMD \geq 3$  with all sample sizes, with one exception. When SMD is 2 and  $n=50$ , the PB in model A.8 is .010, whereas the PB in model B.8 is .003.

When SMD is 1, model B.5 is equal to the model A.5. the PB in model B.5 for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  is lower than the PB in model A.5. When SMD is 2, the PB in model B.5 for  $\alpha_0^{(1)}$  decreases from .042 to .019 when the sample size increases from 50 to 100 and the difference to the PB in model A.5 decreases from .104 to .053. Other than in model A.5, the PB for  $\alpha_0^{(1)}$  is lower than 0.1 percent when  $n=200$ , and is .019 lower than the PB in model A.5. As in model A.5, the PB in model B.5 is lower than .01 when  $n \geq 200$ .

As in model A.5, the PB for  $\alpha_0^{(2)}$  in model B.5 is lower than .01 when  $SMD \geq 3$  with all sample sizes, with two exceptions. Other than in model A.5, the PB in model B.5 is lower than .01 when SMD is 2 and  $n=50$  or  $100$  and the PB is .016 or .009 lower, respectively, than the PB in model A.5.

### **6.3.2.2. Results of PB for $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$**

As can be seen from Table 6.16, the PB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters is lower than .01 in model A.8 with all  $SMD=1, 2, 3, 4$  or  $5$  and with all sample sizes.

Table 6.16. The PB for  $\alpha_1^{(1)}, \alpha_1^{(2)}$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8		A.5		A.5*		B.8		B.5		C.8	
N	SMD	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$
50	1	.0000	.0000	.0002	.0000	.0021	.0020	.0000	.0000	.0002	.0000	.0003	.0000
100	1	.0002	.0000	.0001	.0001	.0027	.0018	.0002	.0000	.0001	.0001	.0003	.0000
200	1	.0001	.0000	.0000	.0001	.0108	.0014	.0001	.0000	.0000	.0001	.0000	.0001
500	1	.0001	.0000	.0000	.0000	.0016	.0006	.0001	.0000	.0000	.0000	.0000	.0005
1000	1	.0000	.0000	.0000	.0000	.0006	.0005	.0000	.0000	.0000	.0000	.0001	.0000
50	2	.0002	.0000	.0005	.0001	.0024	.0024	.0589	.0087	.0679	.0126	.0003	.0000
100	2	.0002	.0000	.0000	.0001	.0031	.0022	.0195	.0052	.0313	.0071	.0000	.0001
200	2	.0002	.0000	.0000	.0000	.0036	.0025	.0022	.0033	.0106	.0052	.0001	.0001
500	2	.0000	.0000	.0001	.0000	.0008	.0003	.0001	.0025	.0003	.0046	.0000	.0001
1000	2	.0001	.0000	.0001	.0004	.0007	.0006	.0000	.0015	.0000	.0022	.0000	.0000
50	3	.0005	.0000	.0005	.0000	.0016	.0028	.0033	.0011	.0130	.0032	.0005	.0001
100	3	.0000	.0000	.0000	.0000	.0010	.0017	.0004	.0006	.0021	.0012	.0001	.0001
200	3	.0000	.0001	.0001	.0002	.0006	.0009	.0002	.0005	.0001	.0010	.0000	.0000
500	3	.0000	.0000	.0000	.0000	.0002	.0001	.0002	.0006	.0002	.0013	.0000	.0000
1000	3	.0000	.0000	.0000	.0000	.0025	.0004	.0001	.0004	.0003	.0009	.0000	.0000
50	4	.0003	.0000	.0002	.0000	.0001	.0008	.0001	.0000	.0013	.0007	.0003	.0000
100	4	.0000	.0000	.0000	.0000	.0001	.0007	.0003	.0000	.0003	.0004	.0000	.0001
200	4	.0000	.0000	.0000	.0000	.0001	.0066	.0001	.0002	.0001	.0004	.0000	.0000
500	4	.0000	.0000	.0000	.0000	.0002	.0001	.0000	.0001	.0001	.0003	.0000	.0001
1000	4	.0000	.0000	.0000	.0000	.0004	.0002	.0000	.0002	.0000	.0003	.0000	.0000
50	5	.0001	.0000	.0002	.0000	.0000	.0004	.0000	.0000	.0001	.0001	.0002	.0000
100	5	.0000	.0001	.0000	.0000	.0000	.0003	.0001	.0000	.0001	.0000	.0001	.0002
200	5	.0001	.0001	.0000	.0000	.0001	.0002	.0001	.0000	.0000	.0001	.0001	.0001
500	5	.0000	.0001	.0000	.0000	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0001
1000	5	.0000	.0000	.0000	.0000	.0004	.0002	.0000	.0000	.0000	.0000	.0000	.0000

Note. The cell is highlighted in grey when the proportion of bias in the MSE is greater than one percent.

*The effect of reliability on the PB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$*

The effect of reliability is not seen in Table 6.16. As in model A.8, the PB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters is lower than .01 in model A.5 with all SMD=1, 2, 3, 4 or 5 and with all sample sizes.

*The effect of additional measurements on the PB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$*

The effect of additional measurement is not seen in Table 6.16. As in model A.8, the PB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters is lower than .01 in model A.5 with all SMD=1, 2, 3, 4 or 5 and with all sample sizes, with one exception. When SMD is 1 and n=200, the PB for  $\alpha_1^{(1)}$  is .011.

*The effect of model construct on the PB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$*

The effect of model construct is not seen in Table 6.16 when comparing model C.8 and A.8. As in model A.8, the PB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters is lower than .01 in model C.8 with all SMD=1, 2, 3, 4 or 5 and with all sample sizes. Other than in model A.8, the PB for  $\alpha_1^{(1)}$  in model B.8 is greater than .01 in two of the cases; namely, when SMD is 2, the PB is .059 or .020 when n=50 or 100, respectively.

Other than in model A.5, the PB for  $\alpha_1^{(1)}$  in model B.5 is greater than .01 in some of the cases. Namely, when SMD is 2, the PB decreases from .068 to .011 when the sample size increases from 50 to 200. When SMD is 3 and n=50, the PB is .013. Other than in model A.5, the PB for  $\alpha_1^{(2)}$  in model B.5 is greater than .01 in one of the cases; namely, when SMD is 2 and n=50, the PB is .013.

**6.3.2.3. Results of PB for  $\psi_{00}, \psi_{11}$  and  $\psi_{01}$**

The results of the PB for  $\psi_{00}, \psi_{11}$  and  $\psi_{01}$  can be seen in Table 6.17. When SMD is 1, the PB is large for  $\psi_{00}$  parameter in model A.8, and decreases from .61 to .118 when the sample size increases from 50 to 1000. When SMD is 2, the PB decreases from .115 to .024 when the sample size increases from 50 to 100, and is lower than .01 when the sample size is  $n \geq 200$ . When SMD is 3, the PB is lower than .01 with all sample sizes. When SMD is 4, the PB decreases from .013 to .011 when the sample size increases from 50 to 100, and is lower than .01 when  $n \geq 200$ . When SMD is 5, the PB decreases from .026 to .014 when the sample size increases from 50 to 100, and is lower than .01 when  $n \geq 200$ .

When SMD is 1, the PB is large for  $\psi_{11}$  parameter in model A.8 and increases from .264 to .317 when the sample size increases from 50 to 1000. When SMD is 2, the PB is also large, and decreases from .186 to .031 when the sample size increases from 50 to 1000. When SMD is 3, the PB decreases from .081 to .014

when the sample size increases from 50 to 200, and is lower than .01 when the sample size is  $n \geq 500$ . When SMD is 4, the PB decreases from .042 to .016 when the sample size increases from 50 to 100, and is lower than .01 when the sample size is  $n \geq 200$ . When SMD is 5, the PB decreases from .034 to .013 when the sample size increases from 50 to 100, and is lower than .01 when  $n \geq 200$ .

When SMD is 1, the PB for  $\psi_{01}$  parameter in model A.8 is lower than .01 when SMD is 1, 2, 3, 4, or 5 with all sample sizes, except with two of the cases. Namely, when SMD is 1, the PB is .013 when the sample size is  $n=50$  or 100.

*The effect of reliability on the PB for  $\psi_{00}, \psi_{11}$  and  $\psi_{01}$*

The effect of reliability on the PB for  $\psi_{00}$  is clear when SMD is 1 or 2. When SMD is 1, the PB in model A.5 decreases from .538 to .222 when the sample size increases from 50 to 1000. The PB is lower in model A.5 when the sample size is 50 or 100, and the difference to the PB in A.8 model is .072 or .031, respectively. The PB is greater in model A.5 when the sample size is  $n \geq 200$ , and the difference to the PB in model A.8 increases from .023 to .103 percent when the sample size increases from 200 to 1000. When SMD is 2, the PB in model A.5 decreases from .172 to .068 when the sample size increases from 50 to 200, and the difference to the PB in model A.8 decreases from .057 to .044, respectively. Other than in model A.8, the PB in model A.5 for  $\psi_{00}$ , is greater than .01, and is .013 greater than the PB in model A.8. As in model A.8, the PB in model A.5 for  $\psi_{00}$  is lower than .01 when  $n \geq 500$ . When SMD is 3, 4 or 5, the PB in model A.5 is lower than .01 with all sample sizes, except with two cases. When SMD is 3 and  $n=50$ , the PB in model A.5 is .013, and is .005 greater than the PB in model A.8. When SMD is 5, the PB in model A.5 is .012, and is .015 lower than the PB in model A.8. Other than in model A.8, the PB in model A.5 is lower than .01 when SMD is 4 and sample sizes are 50 or 100, or when SMD is 5 and the sample size is 100.

The effect of reliability on the PB for  $\psi_{11}$  is weak when SMD is 1 or 2. When SMD is 1, the PB in model A.5 increases from .230 to .308 percent when the sample size increases from 50 to 1000. The PB is lower in model A.5, and the difference decreases from .034 to .009, respectively. When SMD is 2, the PB in model A.5 decreases from .170 to .067 when the sample size increases from 50 to 1000. The PB in model A.8 is .016 greater than the PB in model A.5 when  $n=50$ . The PB in model A.5 is greater than the PB in model A.8 when  $n \geq 100$ . When



the sample size increases from 100 to 500, the difference increases from .008 to .043 and is .035 when n=1000.

Table 6.17. The PB for  $\psi_{00}, \psi_{11}, \psi_{01}$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8			A.5			A.5*		B.8			B.5			C.8		
N	SMD	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$
50	1	.6099	.2641	.0097	.5378	.2298	.0451	.4754	.2378	.6099	.2641	.0097	.5378	.2298	.0451	.6467	.2607	.0187
100	1	.5228	.2832	.0132	.4921	.2557	.0517	.3947	.2051	.5228	.2832	.0132	.4921	.2557	.0517	.5763	.2759	.0169
200	1	.4076	.3042	.0127	.4303	.2850	.0541	.3104	.1698	.4076	.3042	.0127	.4303	.2850	.0541	.4870	.3018	.0194
500	1	.2370	.3164	.0085	.3166	.3053	.0537	.1820	.1354	.2370	.3164	.0085	.3166	.3053	.0537	.3577	.3150	.0236
1000	1	.1184	.3168	.0068	.2215	.3080	.0460	.0928	.1202	.1184	.3168	.0068	.2215	.3080	.0460	.2394	.3264	.0180
50	2	.1153	.1859	.0050	.1724	.1700	.0287	.0905	.1500	.1481	.1109	.0096	.1683	.1528	.0247	.2211	.2056	.0110
100	2	.0242	.1645	.0052	.0677	.1720	.0230	.0166	.1056	.0941	.0400	.0305	.1308	.0969	.0514	.0831	.2013	.0085
200	2	.0001	.1406	.0031	.0134	.1623	.0200	.0008	.0554	.0454	.0095	.0377	.0824	.0482	.0674	.0113	.1855	.0056
500	2	.0003	.0720	.0007	.0000	.1152	.0084	.0003	.0187	.0156	.0021	.0194	.0395	.0149	.0584	.0002	.1344	.0049
1000	2	.0000	.0313	.0002	.0003	.0665	.0035	.0001	.0064	.0077	.0015	.0057	.0180	.0055	.0333	.0008	.0756	.0027
50	3	.0082	.0810	.0027	.0129	.0828	.0101	.0046	.0701	.0662	.0112	.0135	.0854	.0345	.0312	.0161	.1065	.0044
100	3	.0028	.0376	.0013	.0014	.0505	.0057	.0020	.0308	.0334	.0019	.0082	.0514	.0082	.0290	.0015	.0632	.0029
200	3	.0022	.0144	.0010	.0002	.0220	.0015	.0020	.0115	.0134	.0012	.0022	.0224	.0025	.0149	.0007	.0275	.0009
500	3	.0026	.0042	.0001	.0012	.0057	.0004	.0017	.0045	.0045	.0013	.0002	.0072	.0018	.0035	.0018	.0088	.0005
1000	3	.0016	.0013	.0000	.0009	.0017	.0001	.0016	.0000	.0029	.0009	.0000	.0041	.0013	.0008	.0014	.0041	.0005
50	4	.0129	.0416	.0017	.0059	.0370	.0043	.0086	.0370	.0436	.0103	.0039	.0449	.0107	.0159	.0079	.0489	.0028
100	4	.0110	.0157	.0006	.0044	.0138	.0021	.0059	.0177	.0213	.0055	.0013	.0213	.0035	.0077	.0067	.0188	.0014
200	4	.0054	.0063	.0005	.0023	.0049	.0006	.0035	.0072	.0079	.0023	.0005	.0080	.0019	.0026	.0041	.0070	.0004
500	4	.0027	.0020	.0000	.0015	.0013	.0002	.0020	.0041	.0033	.0010	.0000	.0034	.0010	.0006	.0025	.0029	.0005
1000	4	.0017	.0008	.0000	.0009	.0004	.0000	.0018	.0016	.0022	.0003	.0000	.0022	.0005	.0002	.0016	.0013	.0006
50	5	.0261	.0339	.0013	.0116	.0252	.0035	.0174	.0300	.0388	.0192	.0015	.0308	.0107	.0078	.0186	.0338	.0029
100	5	.0143	.0125	.0004	.0084	.0086	.0019	.0081	.0152	.0195	.0080	.0006	.0147	.0045	.0031	.0122	.0125	.0019
200	5	.0054	.0048	.0003	.0033	.0033	.0007	.0047	.0060	.0070	.0029	.0003	.0058	.0020	.0012	.0045	.0048	.0007
500	5	.0023	.0016	.0000	.0000	.0009	.0001	.0024	.0037	.0029	.0012	.0000	.0026	.0006	.0003	.0023	.0020	.0008
1000	5	.0016	.0005	.0001	.0010	.0002	.0000	.0021	.0021	.0018	.0005	.0000	.0016	.0002	.0001	.0016	.0008	.0006

Note. Cell is highlighted when the proportion of bias in the MSE is greater than one percent.

When SMD is 3, the PB in model A.5 decreases and is .083, .051 or .022 when the sample size increases and is 50, 100 or 200, respectively. The PB in model A.5 is .002, .013 or .008 greater, respectively, than the PB in model A.8. As in model

A.8, the PB in model A.5 is lower than .01 when the sample size is  $n \geq 500$ . When SMD is 4, the PB in model A.5 decreases and is .037 or .014 when the sample size increases and is 50 or 100, respectively. The PB in model A.5 is .005 or .004 lower, respectively, than the PB in model A.8. As in model A.8, the PB in model A.5 is lower than .01 when the sample size is  $n \geq 200$ . When SMD is 5, the PB in model A.5 decreases and is .025 or .009 when the sample size increases and is 50 or 100, respectively. The PB in model A.5 is .009 or .004 lower, respectively, than the PB in model A.8. As in model A.8, the PB in model A.5 is lower than .01 when the sample size is  $n \geq 200$ .

The effect of reliability on the PB for  $\psi_{01}$  is clear when SMD is 1 or 2. When SMD is 1, the PB in model A.5 is between .045 and .054 with all sample sizes and is clearly greater than the PB in model A.8. When the PB in model A.8 is lower than .01 when  $SMD \geq 2$  with all sample sizes, the PB in model A.5 is greater than .01 in four cases. Namely, when SMD is 2, the PB decreases from .029 to .020 when the sample size increases from 50 to 200, and when SMD is 3 and  $n=50$ , the PB is .010.

*The effect of additional measurements on the PB for  $\psi_{00}$  and  $\psi_{11}$*

The effect of additional measurements on the PB for  $\psi_{00}$  is clear when SMD is 1 or 2 and the PB for  $\psi_{00}$  in model A.5\* is lower than in model A.5. When SMD is 1, the PB in model A.5\* decreases from .475 to .093 when the sample size increases from 50 to 1000 and the difference to the PB in model A.5 increase from .062 to .129. When SMD is 2, the PB is greater than .01 when  $n=50$  or 100 and is .091 or .017. These values are .082 or .051 greater, respectively, than in model A.5. When SMD is 3 or 4, other than in model A.5, the PB is in model A.5\* lower than .01 with all sample sizes. When SMD is 5, as in model A.5, the PB is greater than .01 when  $n=50$  and the value is .017, which is .006 greater than in model A.5.

The effect of additional measurements on the PB for  $\psi_{11}$  is clear when SMD is 1 or 2 and the PB for  $\psi_{11}$  in model A.5\* is lower than in model A.5, on average. When SMD is 1, the PB in model A.5\* decreases from .238 to .120 when the sample size increases from 50 to 1000, and the difference to the PB in model A.5 increases from equal values to .188, respectively. When SMD is 2, the PB decreases from .150 to .019 when the sample size increases from 50 to 500. The difference to the PB in model A.5 increases from .02 to .10, respectively. As in model A.5, the PB in model A.5\* is lower than .01 when  $n=1000$ . When SMD is 3, the PB in model A.5\* decreases from .070 to .012 when the sample size increases from 50 or 200, and the difference decreases from .013 to .011, respectively. When

SMD is 4, the PB in model A.5\* is .037 or .018 when the sample size is 50 or 100, respectively, and these values of PB are .013 or .020, respectively, lower than in model A.5. When SMD is 5 and when  $n=50$ , the PB in model A.5\* is .030, and is .005 greater than in model A.5. Other than in model A.5, the PB in model A.5\* is greater than .01 and is .015 when  $n=100$ . As in model A.5, the PB in model A.5\* is lower than .01 when  $n \geq 200$

*The effect of model construct on the PB for  $\psi_{00}, \psi_{11}$  and  $\psi_{01}$*

When comparing model C.8 to model A.8, the PB for  $\psi_{00}$  in model C.8 is slightly greater than in model A.8. When SMD is 1, the PB decreases from .647 to .239 when the sample size increases from 50 to 1000, and the difference to the PB in model A.8 increases from .037 to .121, respectively. When SMD is 2, the PB decreases from .221 to .083 when the sample size increases from 50 to 100, and the difference from PB in model A.8 decreases from .106 to .059, respectively. Other than in model A.8, the PB in model C.8 is greater than .01, and is .011 when the sample size is  $n=200$ . As in model A.8, the PB is lower than .01 when  $n \geq 500$ . When SMD is 3, the PB in model A.5 is .016 when  $n=50$ , whereas the PB is lower than .01 in model A.8. Other than in model A.8, the PB in model C.8 is lower than .01 when SMD is 4 also when  $n=50$ . As in model A.8, the PB is greater than .01 when SMD is 5 and the sample size is 50 or 100. The PB is then .019 or .012, respectively, and is .008 or .002 lower than the PB in model A.8.

The PB for  $\psi_{11}$  in model C.8 is slightly greater than in model A.8. When SMD is 1, the PB increases from .261 to .326 when the sample size increases from 50 to 1000 and differs slightly from the PB in model A.8. When SMD is 2, the PB decreases from .206 to .076 when the sample size increases from 50 to 1000 and the difference from the PB in model A.8 varies between .020 and .062. As in model A.8, the PB in model C.8 is greater than .01 when SMD is 3 and  $n=50, 100$  or 200, and the PB decreases from .107 to .028 when the sample size increases from 50 to 200, and the difference from the PB in model A.8 decreases from .025 to .013, respectively. As in model A.8, the PB in model C.8 is greater than .01 when SMD is 4 and  $n=50$  or 100, and the PB decreases from .049 to .019 when the sample size increases from 50 to 100 and the difference from PB in model A.8 decreases from .008 to .003, respectively. As in model A.8, the PB in model C.8 is greater than .01 when SMD is 5 and  $n=50$  or 100, and the PB decreases from .034 to .013 when the sample size increases from 50 to 100 and there are no differences to the PB in model A.8.

Model B.8 is equal to model A.8 when SMD is 1. The PB for  $\psi_{00}$  in model B.8 is slightly greater than in model A.8. When SMD is 2, other than in model A.8, the PB in model B.8 is greater than .01 when  $n=200$  or  $n=500$ . The PB decreases in model B.8 from .148 to .016 when the sample size increases from 50 to 500. When the sample size is 50 or 100, the difference from the PB in model A.8 is .033 or .070, respectively. When the PB in model A.8 is lower than .01 with all sample sizes, the PB in model B.8 decreases from .066 to .013 when the sample size increases from 50 to 200. As in model A.8, the PB in model B.8 is greater than .01 when SMD is 4 or 5 and the sample size is 50 or 100. When SMD is 4, the PB decreases from .045 to .021 when the sample size increases from 50 to 100, and the difference decreases from .032 to .010, respectively. When SMD is 5, the PB decreases from .031 to .015 when the sample size increases from 50 to 100, and the difference decreases from .012 to .0004, respectively.

PB for  $\psi_{11}$  in model B.8 is slightly lower than in model A.8. When SMD is 2, other than in model A.8, the PB in model B.8 is greater than .01 only when  $n=50$  or  $n=100$ . The PB decreases in model B.8 from .111 to .040 when the sample size increases from 50 to 100. These values are .075 and .125 lower than in model A.8. Other than in model A.8, when SMD is 3, 4 or 5, the PB is lower than .01 only when  $n=50$ . These values are then .011, .010 and .019 and are .070, .031 and .015 lower, respectively, than values in model A.8.

As in model A.8, the PB for  $\psi_{01}$  in model B.8 is lower than .01 when SMD is 2, 3, 4 or 5 and with all sample sizes, except with four cases. Three of these exceptions appear when SMD is 2 and the PB in model B.5 is .031, .038 or .019 when the sample size is 100, 200 or 500, respectively. The fourth exception appears when SMD is 3 and the sample size is 50, in which case the PB in model B.8 is .014.

Model B.5 is equal to model A.5 when SMD is 1. The PB for  $\psi_{00}$  in model B.5 is close to the PB in model A.5, but appears to be over .01 more often than in model A.5. When SMD is 2, other than in model A.5, the PB in model B.5 is greater than .01 also when  $n=200$  or  $n=500$ . The PB decreases in model B.5 from .168 to .018 when the sample size increases from 50 to 1000. When the sample size is 50, 100 or 200, the difference from the PB in model A.5 is .004, .063 or .069, respectively. When  $SMD \geq 3$ , the PB in model A.5 is greater than .01 only in two cases. Namely, when SMD is 3 or 5 and the sample size is 50, in which cases the PB in model B.5 is .073 or .019 greater than the PB in model A.5. Addition to these cases, the PB in model B.5 is greater than .01 in many cases. First, when SMD is 3, the PB in model B.5 decreases from .085 to .022 when the sample size increases from 50 to 200. Secondly, when SMD is 4, the PB decreases from .045 to .021

when the sample size increases from 50 to 100. Thirdly, when SMD is 5, the PB in model B.5 decreases from .031 to .015 when the sample size increases from 50 to 100.

The PB for  $\psi_{11}$  in model B.5 is lower than in model A.5. When SMD is 2, other than in model A.5, the PB in model B.5 is lower than .01 when  $n=1000$ . The PB decreases in model B.5 from .153 to .015 when the sample size increases from 50 to 500. These values, when  $n=50, 100, 200$  or  $500$ , are .017, .075, .114, .100 lower, respectively, than in model A.5. Other than in model A.5, when SMD is 3, 4 or 5, the PB is greater than .01 only when  $n=50$ . These values are .035, .011 or .011, respectively, and are .047, .031 or .023 lower, respectively, than the values in model A.5.

The PB for  $\psi_{01}$  in model B.5 is slightly greater than in model A.5. When SMD is 2, other than in model A.5, the PB in model B.5 is greater than .01 also when  $n=500$  or  $n=1000$ . The PB increases in model B.5 from .025 to .067 when the sample size increases from 50 to 200 and decreases after that to .033 when the sample size increases to 1000. These values, when  $n=100$  or  $200$ , are .028 or .054 greater, respectively, than in model A.5. Other than in model A.5, when SMD is 3, the PB is lower than .01 in addition to the case when  $n=50$ , also when  $n=100$  or  $200$ . The PB in model B.5 decreases from .031 to .015. When  $n=50$ , the PB is .021 greater than in model A.5. When the PB in model A.5 is lower than .01 with all sample sizes when SMD is 4 or 5, in model B.5 there is one difference. Namely, when SMD is 4 and the sample size is 50, the PB is .016.

#### **6.3.2.4. Results of PB for $\theta_1, \theta_2, \theta_3$ and $\theta_4$**

As can be seen from Table 6.18, the PBs for  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameters are lower than .01 in model A.8 with all SMD=1, 2, 3, 4 or 5 and with all sample sizes.

##### *The effect of reliability on the PB for $\theta_1, \theta_2, \theta_3$ and $\theta_4$*

The effect of reliability is not evident in Table 6.18. As in model A.8, the PBs for  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameters are lower than .01 in model A.5 with all SMD=1, 2, 3, 4 or 5 and with all sample sizes.

Table 6.18. The PB for  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8				A.5				A.5*				B.8				B.5				C.8			
n	SMD	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_3$	$\theta_5$	$\theta_7$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
50	1	.0001	.0007	.0000	.0000	.0001	.0006	.0000	.0000	.0030	.0003	.0003	.0001	.0001	.0007	.0000	.0000	.0001	.0006	.0000	.0000	.0001	.0006	.0002	.0000
100	1	.0001	.0005	.0001	.0001	.0002	.0003	.0000	.0002	.0019	.0000	.0004	.0002	.0001	.0005	.0001	.0001	.0002	.0003	.0000	.0002	.0002	.0005	.0001	.0002
200	1	.0001	.0004	.0001	.0003	.0001	.0002	.0001	.0006	.0020	.0000	.0000	.0002	.0001	.0004	.0001	.0003	.0001	.0002	.0001	.0006	.0001	.0003	.0002	.0004
500	1	.0002	.0000	.0002	.0001	.0001	.0000	.0002	.0003	.0003	.0002	.0000	.0000	.0002	.0000	.0002	.0001	.0001	.0000	.0002	.0003	.0001	.0000	.0009	.0002
1000	1	.0003	.0001	.0001	.0002	.0001	.0000	.0001	.0004	.0000	.0000	.0001	.0003	.0003	.0001	.0001	.0002	.0001	.0000	.0001	.0004	.0003	.0000	.0001	.0003
50	2	.0003	.0008	.0000	.0000	.0000	.0009	.0001	.0000	.0010	.0003	.0002	.0000	.0002	.0011	.0000	.0001	.0004	.0008	.0000	.0001	.0001	.0006	.0002	.0000
100	2	.0003	.0006	.0001	.0001	.0003	.0005	.0000	.0000	.0006	.0000	.0001	.0001	.0001	.0005	.0001	.0001	.0002	.0005	.0000	.0003	.0003	.0006	.0001	.0002
200	2	.0003	.0006	.0002	.0003	.0002	.0004	.0001	.0007	.0009	.0000	.0000	.0001	.0001	.0004	.0001	.0003	.0001	.0003	.0001	.0007	.0003	.0005	.0002	.0004
500	2	.0002	.0000	.0002	.0001	.0001	.0000	.0002	.0003	.0001	.0001	.0000	.0000	.0001	.0000	.0002	.0001	.0000	.0000	.0001	.0003	.0003	.0000	.0002	.0002
1000	2	.0003	.0001	.0001	.0002	.0001	.0000	.0001	.0004	.0000	.0000	.0001	.0002	.0002	.0000	.0002	.0001	.0001	.0000	.0001	.0003	.0003	.0001	.0001	.0003
50	3	.0001	.0004	.0001	.0000	.0005	.0006	.0002	.0000	.0002	.0002	.0002	.0000	.0002	.0010	.0000	.0001	.0003	.0009	.0001	.0000	.0002	.0006	.0002	.0001
100	3	.0001	.0004	.0001	.0001	.0003	.0004	.0000	.0004	.0003	.0000	.0001	.0001	.0001	.0006	.0001	.0002	.0002	.0006	.0000	.0006	.0002	.0005	.0001	.0002
200	3	.0003	.0004	.0002	.0003	.0003	.0004	.0001	.0007	.0009	.0000	.0000	.0001	.0001	.0003	.0001	.0005	.0000	.0003	.0000	.0010	.0004	.0005	.0002	.0004
500	3	.0002	.0000	.0002	.0001	.0001	.0000	.0001	.0003	.0000	.0001	.0000	.0000	.0001	.0000	.0001	.0002	.0000	.0000	.0001	.0004	.0002	.0000	.0002	.0001
1000	3	.0003	.0001	.0001	.0002	.0001	.0000	.0001	.0004	.0000	.0000	.0001	.0002	.0002	.0000	.0002	.0001	.0001	.0000	.0001	.0002	.0004	.0001	.0001	.0003
50	4	.0001	.0004	.0000	.0000	.0003	.0005	.0001	.0000	.0003	.0001	.0002	.0000	.0001	.0007	.0000	.0000	.0003	.0008	.0001	.0000	.0001	.0004	.0001	.0000
100	4	.0001	.0003	.0002	.0001	.0002	.0004	.0000	.0003	.0005	.0000	.0001	.0001	.0001	.0006	.0001	.0002	.0002	.0006	.0000	.0007	.0001	.0003	.0001	.0002
200	4	.0001	.0003	.0002	.0002	.0002	.0003	.0001	.0007	.0012	.0000	.0000	.0001	.0001	.0003	.0001	.0004	.0000	.0003	.0000	.0010	.0003	.0004	.0002	.0003
500	4	.0002	.0000	.0002	.0001	.0001	.0000	.0001	.0004	.0000	.0001	.0000	.0000	.0001	.0000	.0001	.0002	.0000	.0000	.0001	.0005	.0003	.0000	.0002	.0001
1000	4	.0003	.0001	.0001	.0002	.0002	.0000	.0000	.0004	.0000	.0000	.0001	.0002	.0002	.0000	.0001	.0001	.0000	.0000	.0001	.0003	.0004	.0001	.0001	.0003
50	5	.0001	.0004	.0000	.0000	.0004	.0005	.0001	.0000	.0005	.0001	.0002	.0000	.0000	.0005	.0000	.0000	.0002	.0006	.0001	.0000	.0001	.0004	.0000	.0000
100	5	.0002	.0005	.0002	.0001	.0003	.0004	.0001	.0003	.0007	.0000	.0001	.0002	.0001	.0005	.0001	.0002	.0002	.0005	.0000	.0007	.0001	.0004	.0001	.0002
200	5	.0002	.0003	.0002	.0002	.0002	.0003	.0001	.0006	.0013	.0000	.0000	.0001	.0001	.0003	.0001	.0004	.0000	.0003	.0000	.0009	.0002	.0003	.0002	.0003
500	5	.0002	.0000	.0002	.0001	.0002	.0000	.0001	.0004	.0000	.0001	.0000	.0000	.0001	.0000	.0002	.0002	.0000	.0000	.0001	.0005	.0002	.0000	.0002	.0001
1000	5	.0003	.0000	.0001	.0002	.0002	.0000	.0000	.0005	.0000	.0000	.0001	.0002	.0002	.0000	.0001	.0002	.0000	.0000	.0001	.0003	.0003	.0000	.0001	.0003

*The effect of additional measurements on the PB for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$*

The effect of additional measurement is not evident in Table 6.18. As in model A.5, the PBs for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters are lower than .01 in model A.5\* with all SMD=1, 2, 3, 4 or 5 and with all sample sizes with one exception.

*The effect of model construct on the PB for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$*

The effect of model construct is not evident in Table 6.18. As in model A.8, the PBs for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters are lower than .01 in model C.8 with all SMD=1, 2, 3, 4 or 5 and with all sample sizes. As in model A.8, the PBs for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters are lower than .01 in model B.8 with all SMD=1, 2, 3, 4 or 5 and with all sample sizes. As in model A.5, the PBs for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters are lower than .01 in model B.5 with all SMD=1, 2, 3, 4 or 5 and with all sample sizes.

### **6.3.2.5. Summary of the results of PB**

Parameter estimates are unbiased for  $\alpha_1^{(1)}$ ,  $\alpha_1^{(2)}$  and for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters with all SMD and with all sample sizes. For these parameters, the MSE is then approximately the size of standard error. Instead, for  $\alpha_0^{(1)}$ ,  $\alpha_0^{(2)}$ ,  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters, the parameter estimates are biased when SMD is 1. When SMD is 2, the bias for these parameters is very small, and decreases strongly when the sample size increases. The above results are similar in all models A.8, A.5, A.5\*, B.8, B.5 and C.8.

When SMD is 1, the PB decreases strongly when the sample size increases from 50 to 1000. Even when the sample size is 1000, the PB is greater than .05 for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters, and greater than .09 for  $\psi_{00}$  and  $\psi_{11}$  parameters, in all tested model A.8 – C.8.

When SMD is 2, the PB is greater than .01 for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters in model A.8 when the sample size is 50 or 100. When reliability of observed variables decreases, the PB slightly increases, and the PB is greater than .01 when the sample size is 50, 100 or 200 for these parameters. In model C.8, the PB is slightly greater than in model A.8 and the PB is lower than .01 when n=50, 100 or 200, whereas in model B.8 or B.5, the PB is lower than in model A.8 or A.5,

respectively, and the PB is greater than .01 when  $n=50$  or  $100$ . When SMD is 3, the PB is negligible for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters in all models A.8, A.5, .A.5\*, B.8, B.5 and C.8.

When SMD is 2, the PB for  $\psi_{00}$ ,  $\psi_{11}$ ,  $\psi_{01}$  parameters is highest for  $\psi_{00}$  and decreases from .19 to .03 when the sample size increases from 50 to 1000. The PB is largest for that parameter also when SMD is 3, 4 or 5. When the sample size is 50, the PB is .09, .04 or .03 when SMD is 3, 4 or 5, respectively, but decreases strongly.

When reliability of observed variables decreases, the PB increases, on average, for  $\psi_{00}$ ,  $\psi_{11}$ ,  $\psi_{01}$  parameters. As in model A.8, the PB is highest for  $\psi_{00}$ . When SMD is 2, the PB decreases from .17 to .07 when the sample size increases from 50 to 1000. When the sample size is 50, the PB is .08, .04 or .03 when SMD is 3, 4 or 5, respectively, but decreases strongly when the sample size increases.

The PB in model C.8 for to A.8  $\psi_{00}$ ,  $\psi_{11}$ ,  $\psi_{01}$  parameters is slightly greater than in model A.8, and the effect of model construct is weak. Instead, when comparing model B.8 to model A.8, or model B.5 to model A.5, the construct effect is clearly seen. When in models A.8 and A.5, the greatest PB appears for  $\psi_{00}$ , in model B.8 or B.5, the greatest PB appears for  $\psi_{11}$  parameter. From this point of view, the PB is greater than .01 in model A.8 or A.5 as well as in model B.8 or B.5, when SMD is 3 and the sample size is lower than 200 or when SMD is 4 or 5 and the sample size is lower than 100.

### **6.3.3. Results of relative bias of asymptotic standard error (RB)**

Next, the behaviour of standard errors of parameter estimators as a function of sample size with different SMD is examined. As described previously in section 5.6, RB – the relative bias of asymptotic standard error of parameter estimates - is used. When the standard error estimate is downward biased this leads to the bias of p-value described earlier in section 5.6.

#### **6.3.3.1. Results of RB for $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$**

In model A.8, when SMD is 1, the standard error estimates for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters are strongly downward biased and the RB is between .53 - .60 (see Table 6.19). When SMD is 2, the bias of standard error rapidly decreases when the



sample size increases. As can be seen from Figure 6.26, the RB is, on average, greater than .95 for  $\alpha_0^{(1)}$  when  $n \geq 420$ , and for  $\alpha_0^{(2)}$  when  $n \geq 500$ . It is noticeable that the RB increases over 1 when the sample size increases to  $n=1000$ , the RB being 1.03-1.04. This happens also when SMD is 3, 4 or 5, in which cases the greatest RB is 1.03. When SMD is 3, the sample size needed to achieve  $RB=.95$  is for  $\alpha_0^{(1)}$  parameter  $n=70$ , and for  $\alpha_0^{(2)}$  parameter  $n=90$ . The RB is greater than .90 for  $\alpha_0^{(1)}$  when  $n=50$ , and for  $\alpha_0^{(2)}$  when  $n=70$ . When SMD is 4 or 5, the RB is greater than .95 with smallest sample size  $n=50$ . As discussed in section 5.6, when the RB is lower than .90, the p-value for testing the null hypotheses that the parameter value is zero is, for example, .078 instead of nominal .05 level, or .02 instead of .01 level.

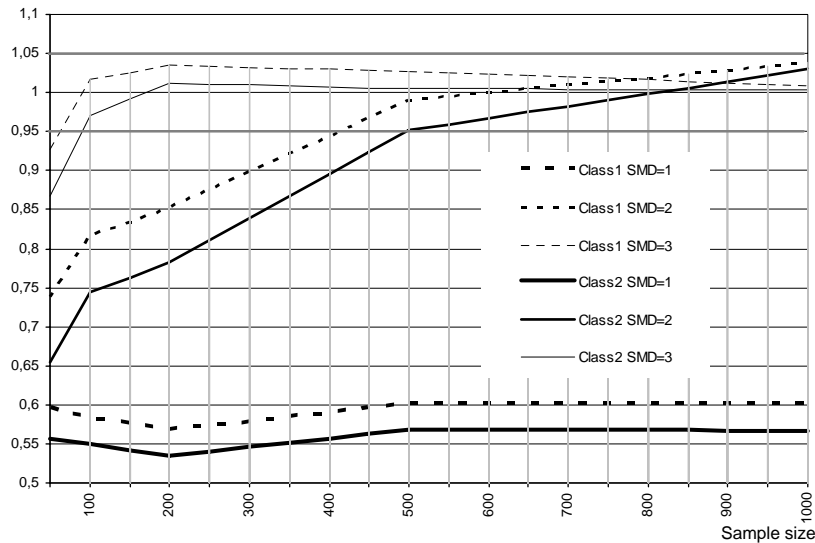


Figure 6.26. The RB in model A.8 for mean  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters when SMD is 1, 2 or 3.

*The effect of reliability on the RB for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

When reliability of observed variables decreases (A.5 vs. A.8), the bias of standard error increases (see Table 6.19.). As in model A.8, when SMD is 1, the standard errors of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters are strongly downward biased also in model A.5, with the RB varying between .52 - .58.

When SMD is 2, the bias of standard error rapidly decreases when the sample size increases. As can be seen and estimated from Table 6.19., the RB for  $\alpha_0^{(1)}$  is greater

than .90 when  $n \geq 500$ , and greater than .95 when  $n \geq 860$ . The standard error of  $\alpha_0^{(2)}$  is more downward biased and the RB for  $\alpha_0^{(2)}$  is only .84 when  $n=1000$ . The required sample sizes to achieve  $RB > .90$  or .95 is about two times larger for  $\alpha_0^{(1)}$  in Model A.5 than in model A.8, whereas for  $\alpha_0^{(2)}$ , a two-times larger sample size is not enough.

Table 6.19. The RB for  $\alpha_0^{(1)}, \alpha_0^{(2)}$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8		A.5		A.5*		B.8		B.5		C.8	
n	SMD	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$
50	1	.5962	.5566	.5650	.5508	0.5807	0.5683	.5962	.5566	.5650	.5508	.5590	.5492
100	1	.5837	.5496	.5687	.5365	0.566	0.5399	.5837	.5496	.5687	.5365	.5542	.5354
200	1	.5676	.5343	.5311	.5219	0.5517	0.5515	.5676	.5343	.5311	.5219	.5437	.5554
500	1	.6015	.5688	.5496	.5415	0.5496	0.5515	.6015	.5688	.5496	.5415	.5700	.5478
1000	1	.6021	.5667	.5803	.5370	0.6079	0.5431	.6021	.5667	.5803	.5370	.5364	.5425
50	2	.7370	.6540	.6825	.6192	0.7185	0.6576	.6832	.6290	.6425	.5790	.6824	.6128
100	2	.8177	.7438	.7343	.6517	0.7333	0.692	.7168	.6763	.6292	.6110	.7211	.6676
200	2	.8520	.7830	.7602	.7010	0.8028	0.7203	.7666	.7458	.6913	.6569	.7931	.7118
500	2	.9908	.9517	.8984	.8074	0.8651	0.8075	.9287	.8577	.7964	.7560	.9256	.8101
1000	2	1.0380	1.0294	.9697	.8362	0.9585	0.9817	1.0200	.9973	.9391	.8430	.9687	.8717
50	3	.9269	.8666	.8588	.7950	0.9235	0.8287	.8646	.8047	.7590	.6924	.8638	.7820
100	3	1.0164	.9707	.9413	.8585	0.9538	0.9281	.9668	.9320	.8406	.7877	.9441	.8608
200	3	1.0348	1.0118	1.0131	.9756	0.9932	0.9817	1.0111	.9989	.9448	.8988	1.0130	.9633
500	3	1.0270	1.0049	1.0314	1.0117	1.0041	1.0027	1.0227	1.0143	1.0204	1.0051	1.0307	1.0119
1000	3	1.0088	1.0035	1.0167	1.0102	1.0059	1.002	1.0020	1.0013	1.0156	1.0060	1.0120	1.0091
50	4	1.0077	.9614	.9769	.9294	0.989	0.8511	.9712	.9286	.8966	.8493	.9899	.9182
100	4	1.0172	.9935	1.0164	.9874	0.9916	0.9736	1.0078	.9874	.9897	.9480	1.0311	.9882
200	4	1.0152	.9987	1.0189	1.0033	0.9944	0.98	1.0050	1.0027	1.0167	1.0087	1.0199	.9994
500	4	1.0182	1.0020	1.0200	1.0046	1.0013	0.9966	1.0000	1.0033	1.0149	1.0080	1.0196	1.0018
1000	4	1.0021	1.0029	1.0066	1.0010	1.0018	0.9952	1.0000	.9985	1.0000	.9989	1.0020	1.0000
50	5	.9895	.9580	.9933	.9673	0.9785	0.944	.9829	.9574	.9686	.9302	1.0019	.9666
100	5	1.0000	.9811	1.0085	.9838	0.984	0.9625	.9978	.9812	1.0140	.9873	1.0117	.9923
200	5	1.0051	.9917	1.0081	.9924	0.9912	0.9746	1.0010	.9934	1.0058	1.0000	1.0107	.9961
500	5	1.0147	1.0011	1.0143	1.0017	0.9972	0.9905	1.0100	1.0000	1.0107	1.0027	1.0172	1.0031
1000	5	1.0000	1.0000	1.0036	1.0075	0.9961	0.9919	1.0000	.9967	.9981	.9974	1.0000	.9985

When SMD is 3 (see Figure 6.27), in model A.5, the RB for  $\alpha_0^{(1)}$  is greater than .90 when  $n \geq 80$ , and greater than .95 when  $n \geq 120$ . These sample sizes are 1.6 – 1.7 times larger than in model A.8. For  $\alpha_0^{(2)}$ , the RB is greater than .90 when  $n \geq 130$ , and greater than .95 when  $n \geq 180$ . These sample sizes are 1.9 – 2.0 times larger than in model A.8.

As in model A.8, the RB is greater than .95 for  $\alpha_0^{(1)}$  when SMD is 4 or 5, and for  $\alpha_0^{(2)}$  when SMD is 5. When SMD is 4, the RB for  $\alpha_0^{(2)}$  is .93 when  $n=50$  and greater than .95 when  $n \geq 70$ .

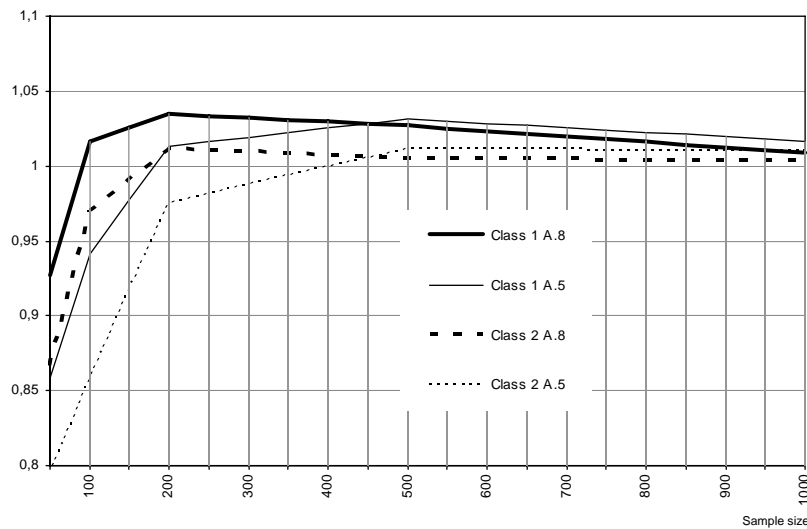


Figure 6.27. The RB for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters in models A.8 and A.5 when SMD is 3.

*The effect of additional measurements on the RB for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

In model A.5\*, the bias of standard error estimates slightly change when compared to the model A.5 (see Table 6.19). As in model A.5, when SMD is 1, the standard errors of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are strongly downward biased also in model A.5\*, in which the RB varies between .54 - .61.

When SMD is 2, the bias of standard error rapidly decreases when the sample size increases. As can be seen from Figure 6.28., the RB for  $\alpha_0^{(1)}$  is greater than .90 when  $n \geq 680$ , and greater than .95 when  $n \geq 950$ . These sample sizes in model A.5\* are 180 and 90 greater, respectively, than in model A.5. The RB for  $\alpha_0^{(2)}$  in

model A.5\* is greater than .90 when  $n \geq 770$ , and greater than .95 when  $n \geq 920$ , whereas in model A.5, the RB is only .84 when  $n=1000$ .

When SMD is 3 (see Figure 6.28), the RB for  $\alpha_0^{(1)}$  in model A.5\* is greater than .90 when  $n \geq 50$ , and greater than .95 when  $n \geq 95$ . In model A.5, these sample sizes were 25 or over 30 larger, respectively, than in model A.5\*. For  $\alpha_0^{(2)}$ , the RB is greater than .90 when  $n \geq 85$ , and greater than .95 when  $n \geq 140$ . In model A.5, these sample sizes were 40 - 45 larger than in model A.5.

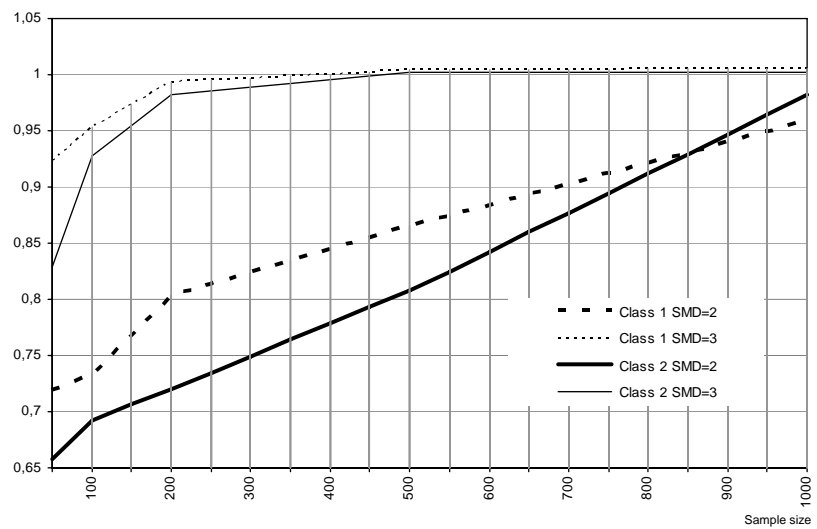


Figure 6.28. The RB in model A.5\* for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters when SMD is 2 or 3.

As in model A.5, the RB for  $\alpha_0^{(1)}$  in model A.5\* is greater than .95 when SMD is 4 or 5. When SMD is 4, the RB for  $\alpha_0^{(2)}$  is only .85 when  $n=50$ , and is greater than .90 when  $n \geq 70$ , and greater than .95 when  $n \geq 90$ . When SMD is 5, the RB for  $\alpha_0^{(2)}$  is .94 when  $n=50$ , and greater than .95 when the sample size is approximately  $n \geq 65$ .

*The effect of model construct on the RB for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

To examine the effect of the construct on standard error estimates, model C.8 is first compared with model A.8. Model C.8 has, on average, clearly more downward biased standard errors of parameter estimates than model A.8. As in

model A.8, when SMD is 1, standard errors of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are strongly downward biased also in model C.8, with the RB varying in model C.8 between .54 - .57.

In the case when SMD is 2, the bias of standard error rapidly decreases when the sample size increases. As can be seen from Figure 6.29., the RB for  $\alpha_0^{(1)}$  is greater than .90 when  $n \geq 440$ , and greater than .95 when  $n \geq 780$ . These sample sizes for  $\alpha_0^{(1)}$  are 120 or 360 greater, respectively, than in model A.8. For  $\alpha_0^{(2)}$ , the RB is only .87 when  $n=1000$ .

As in model A.8, when SMD is 3, 4 or 5, standard error is slightly biased upward, also in model C.8, in which the greatest RB is 1.03. When SMD is 3, the RB for  $\alpha_0^{(1)}$  is greater than .90 when  $n \geq 75$ , and for  $\alpha_0^{(2)}$  when  $n \geq 140$ . These sample sizes for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters are over 25 or 70 greater, respectively, than in model A.8. The RB is greater than .95 for  $\alpha_0^{(1)}$  when  $n \geq 110$ , and for  $\alpha_0^{(2)}$  when  $n \geq 180$ . These sample sizes for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters are 40 or 90 greater, respectively, than in model A.8.

As in model A.8, the RB for  $\alpha_0^{(1)}$  in model C.8 is greater than .95 when SMD is 4 or 5. When SMD is 4, the RB for  $\alpha_0^{(2)}$  is .92 when  $n=50$ , and greater than .95 when  $n \geq 85$ . When SMD is 5, the RB for  $\alpha_0^{(2)}$  is greater than .95 with all sample sizes.

Model B.8 has, on average, clearly more downward biased standard errors of parameter estimates than model A.8. As in model A.8, when SMD is 1, standard errors of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters are strongly downward biased also in model B.8, with the RB varying in model B.8 between .53 - .60.

In the case when SMD is 2, the bias of standard error rapidly decreases when the sample size increases. As can be seen from Figure 6.29. and Figure 6.30., the RB is greater than .90 for  $\alpha_0^{(1)}$  when  $n \geq 450$ , and for  $\alpha_0^{(2)}$  when  $n \geq 650$ . These sample sizes for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are 30 or 150 greater, respectively, than in model A.8. The RB is greater than .95 for  $\alpha_0^{(1)}$  when  $n \geq 620$ , and for  $\alpha_0^{(2)}$  when  $n \geq 830$ . These sample sizes for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are 200 or 330 greater, respectively, than in model A.8.

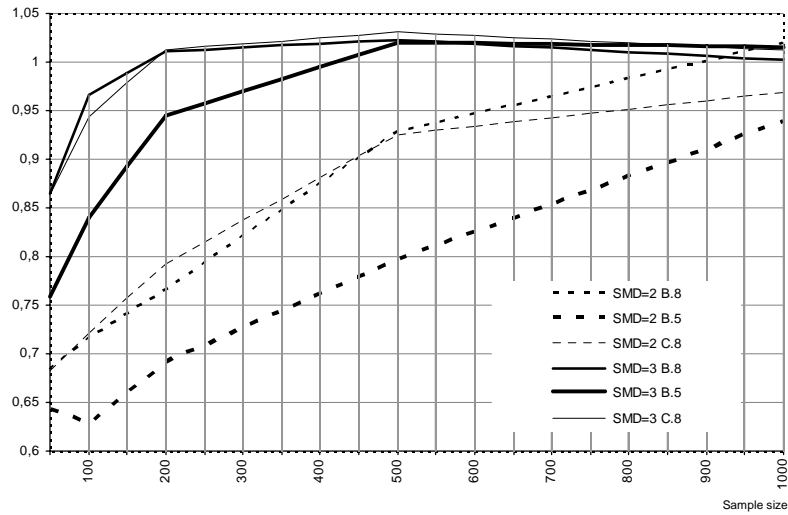


Figure 6.29. The RB for  $\alpha_0^{(1)}$  parameter in model B.8, B.5 and C.5 when SMD is 2 or 3.

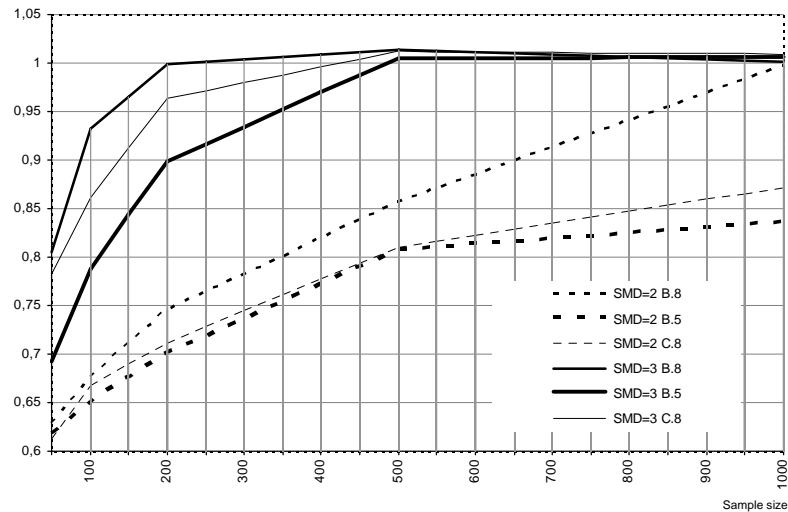


Figure 6.30. The RB for  $\alpha_0^{(2)}$  parameter in model B.8, B.5 and C.8 when SMD is 2 or 3.

When SMD is 3, the RB is greater than .90 for  $\alpha_0^{(1)}$  when  $n \geq 70$  and for  $\alpha_0^{(2)}$  when  $n \geq 85$ . These sample sizes for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are over 20 or 25 greater, respectively, than in model A.8. The RB is greater than .95 for  $\alpha_0^{(1)}$  when  $n \geq 90$

and for  $\alpha_0^{(2)}$  when  $n \geq 135$ . These sample sizes for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are 20 or 45 greater, respectively, than in model A.8.

As in model A.8, the RB for  $\alpha_0^{(1)}$  in model B.8 is greater than .95 when SMD is 4 or 5. When SMD is 4, the RB for  $\alpha_0^{(2)}$  is .93 and greater than .95 when  $n \geq 70$ . When SMD is 5, the RB for  $\alpha_0^{(2)}$  is greater than .95 when  $n \geq 50$ .

Model B.5 has, on average, slightly more downward biased standard errors of parameter estimates than model A.5. As in model A.5, when SMD is 1, the standard errors of  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are strongly downward biased also in model B.5, with the RB varying in model B.5 between .52 - .58.

When SMD is 2, the bias of standard error rapidly decreases when the sample size increases. As can be seen from Figure 6.29, the RB for  $\alpha_0^{(1)}$  is greater than .90 when  $n \geq 860$ . This sample size for  $\alpha_0^{(1)}$  is 360 greater than in model A.5. As in model A.5, the RB is only .84 for  $\alpha_0^{(2)}$  when  $n=1000$ .

When SMD is 3, the RB is greater than .90 for  $\alpha_0^{(1)}$  when  $n \geq 160$ , and for  $\alpha_0^{(2)}$  when  $n \geq 200$ . These sample sizes for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are over 80 or 70 greater, respectively, than in model A.5. The RB is greater than .95 for  $\alpha_0^{(1)}$  when  $n \geq 220$ , and for  $\alpha_0^{(2)}$  when  $n \geq 340$ . These sample size for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are 100 or 160 greater, respectively, than in model A.5.

When SMD is 4, the RB for  $\alpha_0^{(1)}$  is greater than .90 when  $n \geq 50$  and greater than .95 when  $n \geq 80$ . For  $\alpha_0^{(2)}$  parameter, the RB is greater than .90 when  $n \geq 75$  and greater than .95 when  $n \geq 100$ .

As in model A.5, the RB for  $\alpha_0^{(1)}$  in model B.5 is greater than .95 when SMD is 5. For  $\alpha_0^{(2)}$ , the RB is .93 when  $n=50$  and greater than .95 when  $n \geq 70$ .

### **6.3.3.2. Results of RB for $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$**

In the model A.8, the standard error estimates for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters are strongly downward biased when SMD is 1, the RB varying between .51 - .56 (see the next Table 6.20). In the case SMD is 2, the bias of standard error rapidly

decreases when the sample size increases. As can be seen from Figure 6.31, the RB is greater than .90 for  $\alpha_1^{(1)}$  when  $n \geq 490$ , and for  $\alpha_1^{(2)}$  when  $n \geq 620$ . The RB is greater than .95 for  $\alpha_1^{(1)}$  when  $n \geq 700$ , and for  $\alpha_1^{(2)}$  when  $n \geq 820$ . When SMD is 3, the RB is greater than .90 for  $\alpha_1^{(1)}$  when  $n \geq 75$ , and for  $\alpha_1^{(2)}$  when  $n \geq 85$ , and the RB is greater than .95 for  $\alpha_1^{(1)}$  when  $n \geq 100$ , and for  $\alpha_1^{(2)}$  when  $n \geq 130$ . When SMD is 4 or 5, the RB for  $\alpha_1^{(1)}$  is greater than .95 with all sample sizes. When SMD is 4, the RB for  $\alpha_1^{(2)}$  is .93 when  $n=50$ , and is greater than .95 when  $n \geq 65$ . When SMD is 5, the RB for  $\alpha_1^{(2)}$  is .94 when  $n=50$  and is greater than .95 when  $n \geq 60$ .

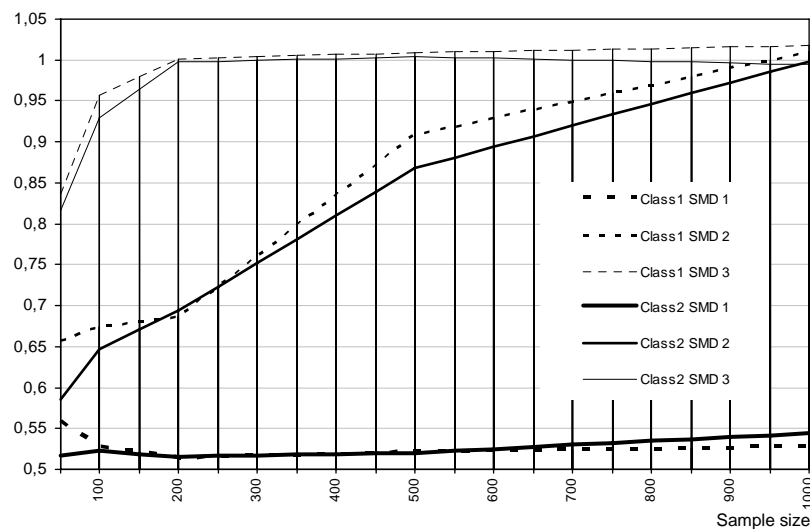


Figure 6.31. The RB for mean  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters in model A.8 when SMD is 1, 2 or 3.

*The effect of reliability on the RB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$*

When reliability of observed variables decreases (A.5 vs. A.8), the bias of standard error increases (see Table 6.20). As in model A.8, when SMD is 1, the standard errors of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters in model A.5 are strongly downward biased, with the RB varying in model A.5 between .50 - .54.

When SMD is 2, the bias of standard error rapidly decreases when the sample size increases. As can be seen from Table 6.20., when  $n=1000$ , the RB is .88 and .86 for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters, respectively.



Table 6.20. The RB for  $\alpha_1^{(1)}, \alpha_1^{(2)}$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8		A.5		A.5*		B.8		B.5		C.8	
n	SMD	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$
50	1	.5597	.5170	.5361	.5068	0.4996	0.5036	.5597	.5170	.5361	.5068	.5557	.5053
100	1	.5278	.5222	.5227	.5148	0.4809	0.4807	.5278	.5222	.5227	.5148	.5145	.5118
200	1	.5139	.5156	.5133	.5014	0.4879	0.4879	.5139	.5156	.5133	.5014	.5117	.4956
500	1	.5212	.5202	.5070	.5031	0.4221	0.4778	.5212	.5202	.5070	.5031	.5170	.5394
1000	1	.5277	.5443	.5406	.5395	0.3883	0.4572	.5277	.5443	.5406	.5395	.4944	.5131
50	2	.6579	.5862	.5969	.5629	0.5926	0.6074	.7052	.6429	.6447	.5943	.6037	.5530
100	2	.6736	.6470	.6002	.5862	0.5896	0.595	.7784	.7087	.6627	.6189	.5924	.5887
200	2	.7438	.6943	.6263	.6331	0.6177	0.6374	.8173	.7417	.7092	.6801	.6570	.6282
500	2	.9092	.8677	.7737	.7439	0.7705	0.6818	.8977	.8739	.8250	.7666	.7427	.7513
1000	2	1.0101	.9987	.8761	.8592	0.946	0.8489	1.0043	.9958	.9410	.8627	.8613	.8668
50	3	.8366	.8162	.7769	.7198	0.8421	0.774	.9232	.8503	.7963	.7312	.7864	.7115
100	3	.9573	.9298	.8387	.7912	0.9281	0.8561	.9857	.9483	.8627	.8023	.8333	.8181
200	3	1.0018	.9978	.9737	.9452	0.9492	0.9602	1.0015	1.0070	.9523	.9118	.9685	.9482
500	3	1.0090	1.0038	1.0214	1.0139	0.995	1.0033	1.0000	.9967	1.0062	.9634	1.0179	.9954
1000	3	1.0173	.9946	1.0141	1.0051	1.000	1.000	.9964	.9791	1.0000	.9927	1.0187	1.0044
50	4	.9555	.9344	.9354	.8856	0.9561	0.9328	.9778	.9432	.9377	.8658	.9287	.8925
100	4	.9884	.9814	.9891	.9680	0.9789	0.9675	.9987	.9845	.9949	.9638	.9974	.9829
200	4	.9836	.9804	.9986	.9914	0.98	0.9862	.9888	.9912	1.0062	1.0016	.9924	.9900
500	4	.9934	.9821	1.0023	.9922	0.992	1.000	.9880	.9840	1.0020	.9895	.9969	.9779
1000	4	1.0095	.9841	1.0100	.9912	0.9887	0.9974	.9915	.9802	.9943	.9796	1.0000	.9855
50	5	.9604	.9428	.9600	.9394	0.9662	0.9331	.9812	.9638	.9754	.9330	.9523	.9490
100	5	.9806	.9678	.9861	.9743	0.9818	0.9641	.9871	.9758	1.0010	.9839	.9886	.9786
200	5	.9810	.9721	.9878	.9790	0.9829	0.9844	.9818	.9820	.9888	.9860	.9820	.9833
500	5	.9932	.9767	.9976	.9832	0.9891	0.9981	.9871	.9781	.9933	.9836	.9904	.9780
1000	5	1.0048	.9834	1.0070	.9857	0.9885	0.9919	.9954	.9783	.9968	.9789	.9954	.9843

When SMD is 3 (see Figure 6.32), in model A.5, the RB for  $\alpha_1^{(1)}$  is greater than .90 when  $n \geq 145$  and greater than .95 when  $n \geq 180$ . These sample sizes for  $\alpha_1^{(1)}$  are 65 or 80 larger, respectively, than in model A.8. For  $\alpha_1^{(2)}$  parameter, the RB is greater than .90 when  $n \geq 170$  and greater than .95 when  $n \geq 220$ . These sample sizes for  $\alpha_1^{(1)}$  are 95 or 90 larger than in model A.8.

Contrary to the results in model A.8, in model A.5, the RB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  is lower than .95 when SMD is 4 or 5 and  $n=50$ . When SMD is 4, the RB for  $\alpha_1^{(1)}$  is greater than .95 when  $n \geq 55$ . When SMD is 4, the RB for  $\alpha_1^{(2)}$  is greater than .90 when  $n \geq 60$ , and greater than .95 for  $\alpha_1^{(2)}$  when  $n \geq 90$ . When SMD is 5, the RB for  $\alpha_1^{(2)}$  is greater than .95 when  $n \geq 65$ .

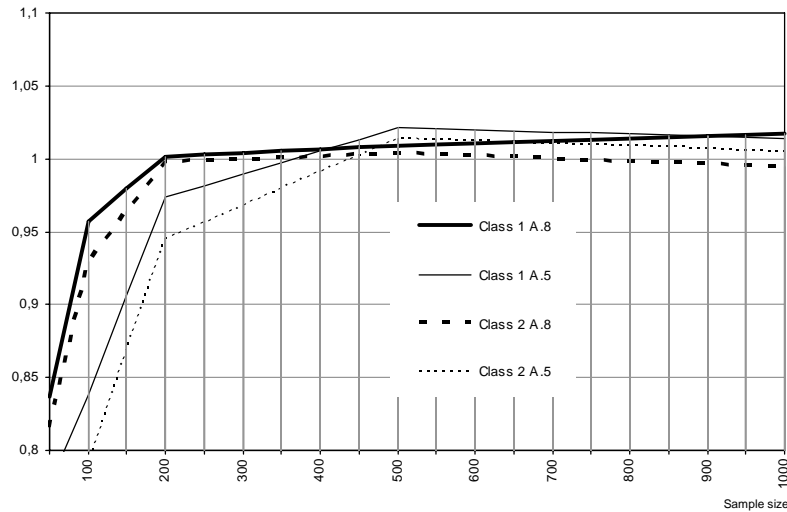


Figure 6.32. The RB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters in model A.8 and A.5 when SMD is 3.

*The effect of additional measurements on the RB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$*

In model A.5\*, the bias of standard error estimates slightly changes compared with model A.5 (see Table 6.20). As in model A.5, when SMD is 1, the standard errors of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters are strongly downward biased also in model A.5\*, with the RB varying in model A.5\* between .54 - .61.

When SMD is 2, the bias of standard error rapidly decreases when the sample size increases. As can be seen from Figure 6.33., the RB for  $\alpha_1^{(1)}$  is greater than .90 when  $n \geq 870$  and is greater than .95 when  $n \geq 1000$ . In model A.5, the RB for  $\alpha_1^{(1)}$  was lower than .90 when  $n=1000$ . The RB for  $\alpha_1^{(2)}$  in model A.5\* is .85 when  $n=1000$ .

When SMD is 3 (see Figure 6.33), in model A.5\*, the RB for  $\alpha_1^{(1)}$  is greater than .90 when  $n \geq 80$  and greater than .95 when  $n \geq 200$ . In model A.5, these sample

sizes were 65 larger or 20 lower, respectively, than in model A.5\*. The RB for  $\alpha_1^{(2)}$  is greater than .90 when  $n \geq 140$  and greater than .95 when  $n \geq 190$ . In model A.5, these sample sizes were 30 larger than in model A.5\*.

Contrary to the results in model A.5, the RB for  $\alpha_1^{(1)}$  in model A.5\* is greater than .95 when SMD is 4 or 5 and  $n=50$ . When SMD is 4 or 5, the RB for  $\alpha_1^{(2)}$  parameter is only .93 when  $n=50$  and greater than .95 when  $n \geq 75$ .

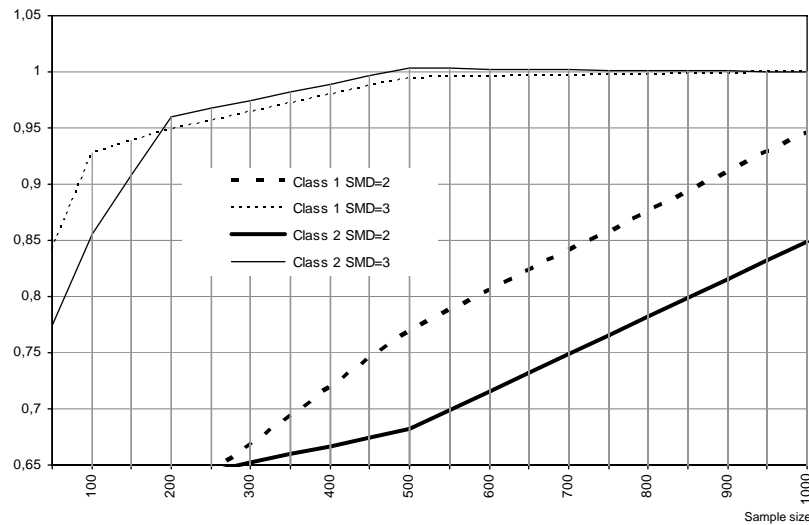


Figure 6.33. The RB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters in model A.5\* when SMD is 2 or 3.

*The effect of model construct on the RB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$*

To examine the effect of the construct on the standard error estimate, model C.8 is first compared with model A.8. Model C.8 has, on average, clearly more downward biased standard errors of parameter estimates than model A.8. As in model A.8, the standard errors of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters are strongly downward biased in model C.8 when SMD is 1, with the RB varying in model C.8 between .49 - .56.

When SMD is 2, the bias of standard error of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters rapidly decreases when the sample size increases (see Figure 6.34 and Figure 6.35). When  $n=1000$ , the RB is only .86 and .87, respectively.

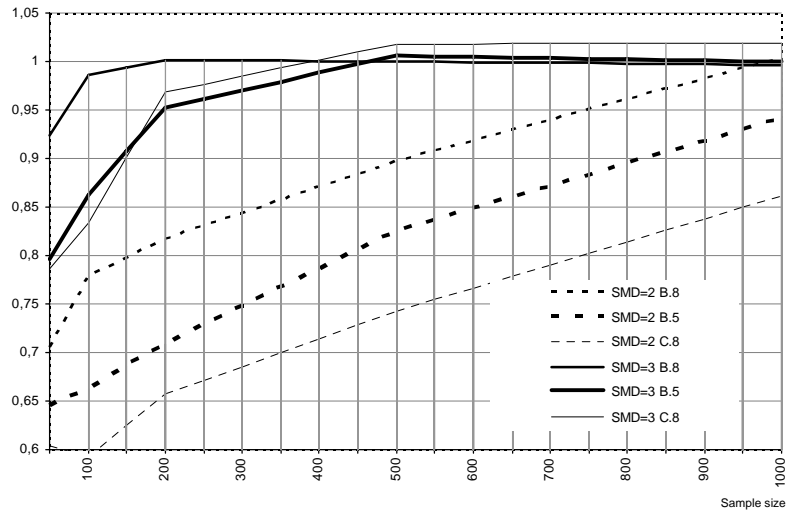


Figure 6.34. The RB for  $\alpha_1^{(1)}$  parameter in model B.8, B.5 and C.8 when SMD is 2 or 3.

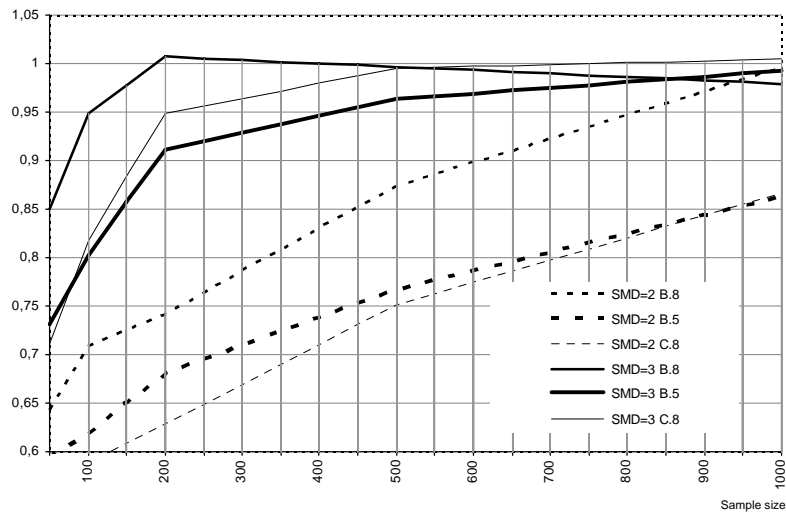


Figure 6.35. The RB for  $\alpha_1^{(2)}$  parameter in model B.8, B.5 and C.8 when SMD is 2 or 3.

Contrary to the results in model A.8, in model C.8 the standard error is slightly biased upward in only a few cases when SMD is 3 or 4 and the greatest RB is 1.02. When SMD is 3, the RB is greater than .90 for  $\alpha_1^{(1)}$  when  $n \geq 150$ , and for  $\alpha_1^{(2)}$  when  $n \geq 165$ . These sample sizes for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  are 75 or 80 greater, respectively, than in model A.8. The RB for  $\alpha_1^{(1)}$  is greater than .95 when  $n \geq 185$ , and for  $\alpha_1^{(2)}$  when  $n \geq 165$ . These sample sizes for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters are 85 or 35 greater, respectively, than in model A.8.

When SMD is 4, the RB for  $\alpha_1^{(1)}$  in model C.8 is .93 when  $n=50$  and greater than .95 when  $n \geq 65$ . The RB for  $\alpha_1^{(2)}$  is .89 when  $n=50$ , greater than .90 when  $n \geq 55$  and greater than .95 when  $n \geq 80$ . When SMD is 5, the RB for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  is .95 when  $n \geq 50$ .

Model B.8 has, on average, slightly lower downward biased standard errors of parameter estimates than model A.8. As in model A.8, in model B.8 the standard errors of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  are strongly downward biased when SMD is 1, the RB varying in model B.8 between .51 - .56.

When SMD is 2, the bias of standard error rapidly decreases when the sample size increases. As can be seen from Figure 6.34, the RB is greater than .90 for  $\alpha_1^{(1)}$  when  $n \geq 510$ , and for  $\alpha_1^{(2)}$  when  $n \geq 600$ . These sample sizes for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  are 10 greater or 20 lower, respectively, than in model A.8. The RB is greater than .95 for  $\alpha_1^{(1)}$  when  $n \geq 750$ , and for  $\alpha_1^{(2)}$  when  $n \geq 810$ . These sample sizes for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  are 50 greater or 10 lower, respectively, than in model A.8.

When SMD is 3, the RB for  $\alpha_1^{(1)}$  is greater than .92 when  $n=50$ . The RB for  $\alpha_1^{(2)}$  is greater than .90 when  $n \geq 75$ . This sample size is 10 smaller than in model A.8. The RB is greater than .95 for  $\alpha_1^{(1)}$  when  $n \geq 70$ , and for  $\alpha_1^{(2)}$  when  $n \geq 100$ . These sample sizes for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are 30 smaller than in model A.8.

As in model A.8, the RB for  $\alpha_1^{(1)}$  in model B.8 is greater than .95 when SMD is 4 or 5 and  $n \geq 50$ . When SMD is 4, the RB for  $\alpha_1^{(2)}$  parameter is .94 and greater than .95 when  $n \geq 60$ . When SMD is 5, the RB for  $\alpha_1^{(2)}$  is greater than .95 when  $n \geq 50$ .

As in model A.5, the standard errors of  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters are strongly downward biased also in model B.5 when SMD is 1, the RB varying in model B.5 between .50 - .54.

When SMD is 2, the bias of standard error in model B.5 rapidly decreases when the sample size increases. As can be seen from Figure 6.34., in model B.5, the RB for  $\alpha_1^{(1)}$  is greater than .90 when  $n \geq 810$ , whereas in model A.5 the RB was .88 when  $n=1000$ . Whereas in model A.5 the RB for  $\alpha_1^{(2)}$  was .94 when  $n=1000$ , in model B.5 this RB is only .86.

When SMD is 3, the RB is greater than .90 for  $\alpha_1^{(1)}$  when  $n \geq 140$ , and for  $\alpha_1^{(2)}$  when  $n \geq 190$ . These sample sizes for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  are only 5 or 10 greater, respectively, than in model A.5. The RB is greater than .95 for  $\alpha_1^{(1)}$  when  $n \geq 195$ , and for  $\alpha_1^{(2)}$  when  $n \geq 425$ . These sample sizes for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  are 25 greater than in model A.5.

When SMD is 4, the RB for  $\alpha_1^{(1)}$  is greater than .95 when  $n \geq 60$ . For  $\alpha_1^{(2)}$ , the RB is greater than .90 when  $n \geq 65$  and greater than .95 when  $n \geq 95$ . When SMD is 5, the RB for  $\alpha_1^{(1)}$  is greater than .97 when  $n \geq 50$ , and for  $\alpha_1^{(2)}$  greater than .95 when  $n \geq 65$ .

### 6.3.3.3. Results of RB for $\psi_{00}$ , $\psi_{11}$ and $\psi_{01}$

In model A.8, the standard error estimates for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters when SMD is 1 are clearly downward biased (see Table 6.21). For  $\psi_{00}$ , the RB increases from .85 to .96 when the sample size increases from  $n=50$  to  $n \geq 500$ ; for  $\psi_{11}$  the RB is stable and lower than .85; and for  $\psi_{01}$  the RB slightly increases from .72 to .76 when the sample size increases from  $n=50$  to  $n=1000$ .

When SMD is 2, the bias of standard error decreases when the sample size increases. As can be seen from Figure 6.36, the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 80$ , for  $\psi_{11}$  when  $n \geq 85$ , and for  $\psi_{01}$  when  $n \geq 340$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 385$ , for  $\psi_{11}$  when  $n \geq 255$ , and for  $\psi_{01}$  when  $n \geq 480$ .

When SMD is 3, the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 60$ , and for  $\psi_{01}$  when  $n \geq 75$ . For  $\psi_{11}$  parameter the RB is .93 when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 130$ , for  $\psi_{11}$  when  $n \geq 75$ , and for  $\psi_{01}$  when  $n \geq 100$ .

When SMD is 4, the RB is .91 for  $\psi_{00}$ , slightly lower than .95 for  $\psi_{11}$ , and .92 for  $\psi_{01}$  when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 85$ , for  $\psi_{11}$  when  $n \geq 65$ , and for  $\psi_{01}$  when  $n \geq 70$ . When SMD is 5, the RB is .94 for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 75$ , for  $\psi_{11}$  when  $n \geq 70$ , and for  $\psi_{01}$  when  $n \geq 65$ .

*The effect of reliability on the RB for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$*

When reliability of observed variables decreases (A.5 vs. A.8), the bias of standard error changes only slightly (see Table 6.21). When SMD is 1, the RB for  $\psi_{00}$  in model A.5 increases from .84 to .95 when the sample size increases from n=50 to n=1000. For  $\psi_{11}$ , the RB is stable and lower than .89, and for  $\psi_{01}$  it slightly increases from .76 to .80 when the sample size increases from n=50 to n=1000.

Table 6.21. The RB for  $\psi_{00}, \psi_{11}, \psi_{01}$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8			A.5			A.5*		B.8			B.5			C.8		
n	SMD	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$
50	1	.8546	.8490	.7186	.8370	.8754	.7632	.7764	.7192	.8546	.8490	.7186	.8370	.8754	.7632	.8447	.8530	.7207
100	1	.9106	.8460	.7226	.8833	.8762	.7723	.8298	.7568	.9106	.8460	.7226	.8833	.8762	.7723	.8925	.8392	.7219
200	1	.9216	.8438	.7248	.8897	.8860	.7707	.8677	.7714	.9216	.8438	.7248	.8897	.8860	.7707	.9387	.8388	.7381
500	1	.9551	.8408	.7458	.9161	.8746	.7746	.8967	.7462	.9551	.8408	.7458	.9161	.8746	.7746	.9749	.8685	.7541
1000	1	.9644	.8409	.7603	.9484	.8762	.8004	.8971	.7304	.9644	.8409	.7603	.9484	.8762	.8004	.9527	.8228	.7511
50	2	.8607	.8725	.7413	.8457	.8845	.7771	.8395	.7952	.8081	.8209	.8297	.8334	.8257	.8379	.8577	.8660	.7355
100	2	.9258	.9136	.7860	.8929	.9033	.7950	.8695	.8553	.8508	.8690	.8523	.8497	.8483	.8554	.9157	.8767	.7624
200	2	.9121	.9390	.8491	.9144	.9363	.8273	.8901	.8858	.9030	.8967	.8830	.8880	.8813	.8992	.9348	.9038	.8173
500	2	.9737	1.0000	.9568	.9683	.9828	.9095	.9201	.9705	.9675	.9562	.9452	.9412	.9404	.9286	.9692	.9468	.9119
1000	2	1.008	1.020	1.026	.9704	1.029	.9564	.9636	1.0049	1.025	1.007	1.027	.9926	.9831	.9892	.9828	1.000	.9494
50	3	.8896	.9320	.8366	.8865	.9324	.8352	.8663	.8976	.9047	.8870	.8534	.8769	.8571	.8556	.8633	.9176	.8192
100	3	.9358	.9716	.9558	.9131	.9615	.8983	.9143	.9644	.9628	.9436	.9358	.9318	.8992	.9146	.9073	.9459	.8883
200	3	.9831	.9905	1.021	.9699	1.000	.9860	.9578	.9909	.9960	.9785	.9953	.9851	.9613	.9587	.9674	1.000	.9966
500	3	.9936	1.000	1.016	.9961	1.000	1.007	.9984	.9964	.9957	1.007	1.005	1.005	1.009	.9987	.9992	1.005	1.010
1000	3	.9871	1.000	1.013	.9979	1.003	1.000	.9901	1.000	.9938	.9949	1.011	1.006	1.007	1.008	.9933	1.000	1.025
50	4	.9147	.9454	.9198	.9049	.9545	.9094	.9219	.9415	.9437	.9161	.9070	.9293	.9099	.8965	.9141	.9480	.8969
100	4	.9635	.9619	.9915	.9530	.9717	.9708	.968	.9685	.9738	.9738	.9849	.9746	.9717	.9609	.9582	.9643	.9980
200	4	.9807	.9840	1.004	.9855	.9927	1.000	.9812	.9881	.9818	.9864	1.002	.9918	1.005	.9956	.9835	.9857	1.0015
500	4	.9783	1.000	.9972	.9809	.9954	.9933	.9892	.9962	.9809	1.000	1.000	.9910	1.000	1.000	.9844	1.000	.9928
1000	4	.9846	1.000	1.008	.9906	1.000	.9981	.9959	.9947	.9814	.9938	1.0084	.9941	.9973	1.000	.9834	1.000	1.018
50	5	.9384	.9448	.9377	.9303	.9569	.9316	.9511	.9294	.9454	.9516	.9530	.9435	.9302	.9219	.9467	.9523	.9481
100	5	.9635	.9595	.9806	.9749	.9672	.9832	.9757	.9657	.9652	.9666	.9888	.9740	.9812	.9849	.9685	.9618	.9964
200	5	.9768	.9839	.9981	.9833	.9912	.9953	.9831	.9902	.9758	.9850	.9980	.9833	.9948	.9971	.9772	.9856	.9948
500	5	.9764	1.000	.9937	.9770	.9977	.9853	.988	.9923	.9780	1.005	1.000	.9857	.9958	.9895	.9782	1.000	.9945
1000	5	.9833	.9928	1.000	.9915	1.000	1.002	1.000	1.000	.9775	1.000	1.005	.9909	.9971	1.000	.9829	.9935	1.008

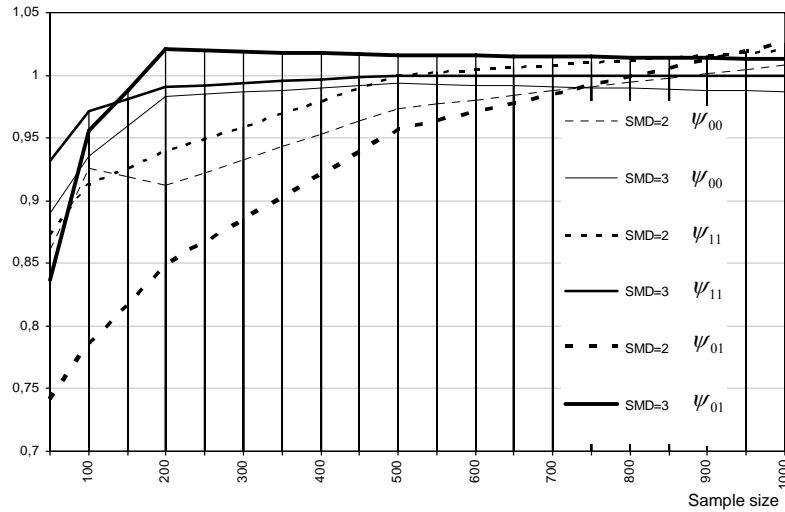


Figure 6.36. The RB for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters in model A.8 when SMD is 2 or 3.

When SMD is 2, the bias of standard error decreases when the sample size increases. As can be seen from Figure 6.37, the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 135$ , for  $\psi_{11}$  when  $n \geq 90$ , and for  $\psi_{01}$  when  $n \geq 465$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 55, 5 and 125 larger, respectively, in model A.5 than in model A.8. The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 400$ , for  $\psi_{11}$  when  $n \geq 290$ , and for  $\psi_{01}$  when  $n \geq 930$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 15, 35 and 455 larger, respectively, in model A.5 than in model A.8.

When SMD is 3, the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 75$ , and for  $\psi_{01}$  when  $n \geq 100$ . These sample sizes for  $\psi_{00}$  and  $\psi_{01}$  are 15 and 25 larger, respectively, in model A.5 than in model A.8. For  $\psi_{11}$ , the RB is .93 when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 165$ , for  $\psi_{11}$  when  $n \geq 80$ , and for  $\psi_{01}$  when  $n \geq 160$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 35, 5 and 60 larger, respectively, in model A.5 than in model A.8.

As in model A.8, when SMD is 4, the RB is high in model A.5 being .91 for  $\psi_{00}$ , .95 for  $\psi_{11}$ , and .91 for  $\psi_{01}$  when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 100$ , and for  $\psi_{01}$  when  $n \geq 85$ . When SMD is 5, the RB is .93 for  $\psi_{00}$ , .96 for  $\psi_{11}$ , and .93 for  $\psi_{01}$  when  $n=50$ . The RB is greater than .95 for both  $\psi_{00}$  and  $\psi_{01}$



when  $n \geq 70$ . This sample size for  $\psi_{00}$  and  $\psi_{01}$  parameters is 10-15 larger in model A.5 than in model A.8.

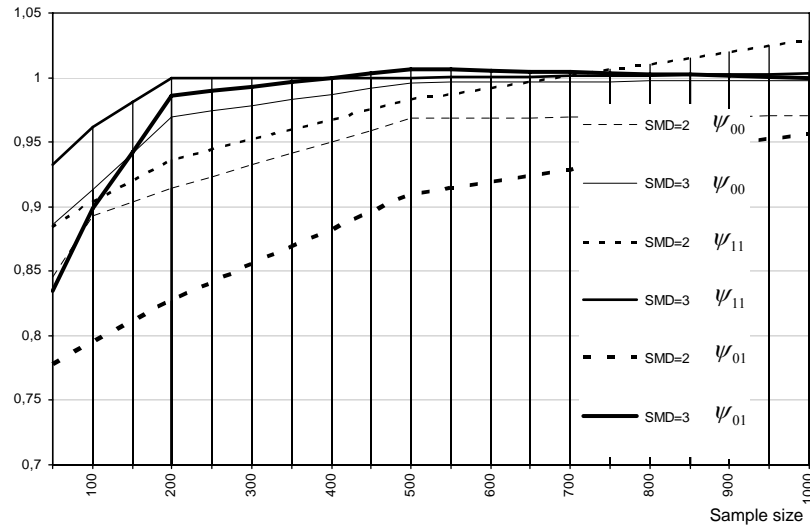


Figure 6.37. The RB for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters in model A.5 when SMD is 2 or 3.

*The effect of additional measurements on the RB for  $\psi_{00}$  and  $\psi_{11}$*

When the reliability of observed variables decreases (A.5 vs. A.8), the bias of standard error changes only slightly (see Table 6.21). When SMD is 1, the RB for  $\psi_{00}$  in model A.5\* increases from .78 to .90 when the sample size increases from  $n=50$  to  $n=1000$ . For  $\psi_{11}$  parameter, the RB varies and is between .72 and .77.

When SMD is 2, the bias of standard error rapidly decreases when the sample size increases. As can be seen from Figure 6.38, the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 300$ , and for  $\psi_{11}$  when  $n \geq 250$ . These sample sizes for  $\psi_{00}$  and  $\psi_{11}$  parameters are 165 and 160 greater, respectively, in model A.5\* than in model A.5. The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 845$  and for  $\psi_{11}$  when  $n \geq 425$ . These sample sizes for  $\psi_{00}$  and  $\psi_{11}$  parameters are 445 and 135 greater, respectively, in model A.5\* than in model A.5.

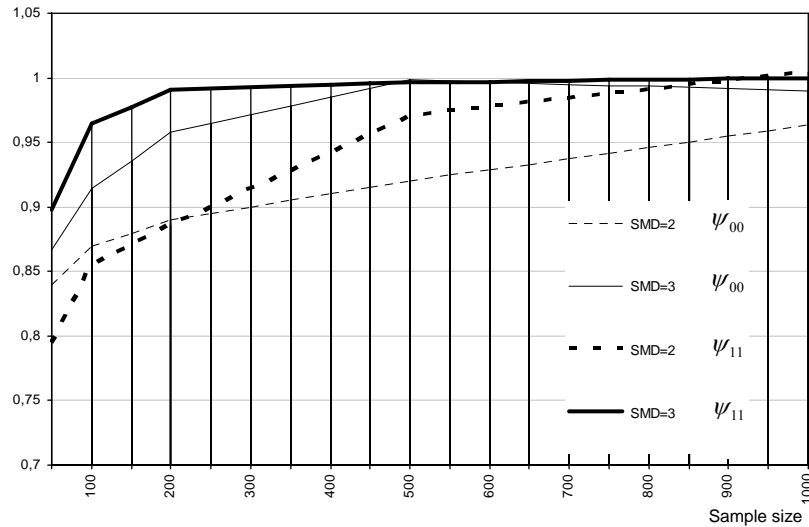


Figure 6.38. The RB for  $\psi_{00}$  and  $\psi_{11}$  parameters in model A.5\* when SMD is 2 or 3.

When SMD is 3, the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 85$ , and for  $\psi_{11}$  when  $n \geq 50$ . These sample sizes for  $\psi_{00}$  are 10 larger in model A.5\* than in model A.5. The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 180$ , and for  $\psi_{11}$  when  $n \geq 90$ . These sample sizes for  $\psi_{00}$  and  $\psi_{11}$  parameters are 15 and 10 larger, respectively, in model A.5\* than in model A.5.

As in model A.5, when SMD is 4, the RBs are high in model A.5\*, being .92 for  $\psi_{00}$  and .94 for  $\psi_{11}$  when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 80$ , and for  $\psi_{11}$  when  $n \geq 65$ . These sample sizes for  $\psi_{00}$  and  $\psi_{11}$  parameters 20 larger in model A.5\* than in model A.5. When SMD is 5, the RB is .95 for  $\psi_{00}$ , and .93 for  $\psi_{11}$ , when  $n=50$ . The RB for  $\psi_{11}$  is greater than .95 when  $n \geq 80$ .

*The effect of model construct on the RB for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$*

To examine the effect of the construct on the standard error estimate model C.8 is first compared to A.8 (see Table 6.21. above). When SMD is 1, the RB for  $\psi_{00}$  in model C.8 increases from .84 to .97, when the sample size increases from  $n=50$  to  $n=500$ , the RB staying a .95 when the sample size increases to  $n=1000$ . For  $\psi_{11}$ , the RB varies and is between .82 - .87, and for  $\psi_{01}$  the RB slightly increases from .72 to .75 when the sample size increases from  $n=50$  to  $n=1000$ .

When SMD is 2, the bias of standard error decreases when the sample size increases. As can be seen from Figure 6.39., the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 85$ , for  $\psi_{11}$  when  $n \geq 185$ , and for  $\psi_{01}$  when  $n \geq 460$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 5, 100 and 120 larger, respectively, in model C.8 than in model A.8. The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 335$ , for  $\psi_{11}$  when  $n \geq 530$ , and for  $\psi_{01}$  when  $n \geq 1000$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 50 lower, 275 and over 500 larger, respectively, in model C.8 than in model A.8.

When SMD is 3, the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 90$ , and for  $\psi_{01}$  when  $n \geq 110$ . These sample sizes for  $\psi_{00}$  and  $\psi_{01}$  are 30 and 10 larger, respectively, in model C.8 than in model A.8. For  $\psi_{11}$  parameter, the RB is .92 when the sample size is  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 170$ , for  $\psi_{11}$  when  $n \geq 105$ , and for  $\psi_{01}$  when  $n \geq 155$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 10, 30 and 55 larger, respectively, in model C.8 than in model A.8.

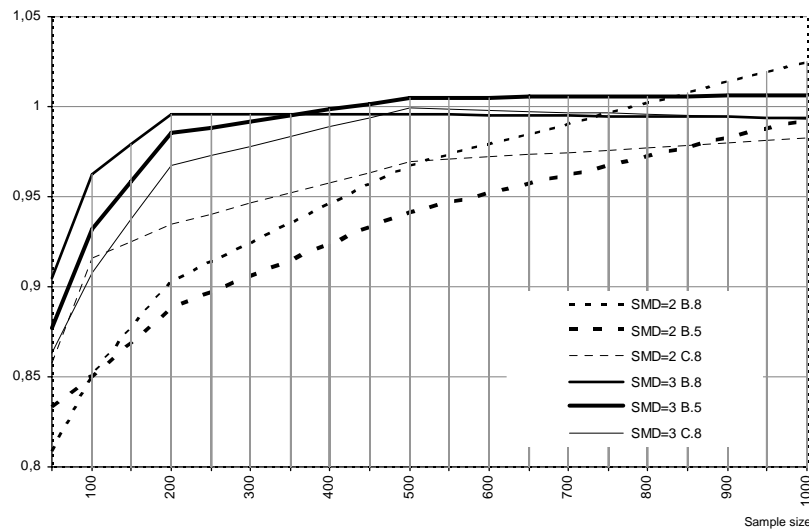


Figure 6.39. The RB for  $\psi_{00}$  parameter in model B.8, B.5 and C.5 when SMD is 2 or 3.

As in model A.8, when SMD is 4, the RBs are high in model C.8, being .91 for  $\psi_{00}$ , .95 for  $\psi_{11}$ , and .90 for  $\psi_{01}$  when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 90$ , for  $\psi_{11}$  when  $n \geq 55$ , and for  $\psi_{01}$  when  $n \geq 75$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 5 larger, 10 smaller and 5 larger,

respectively, in model C.8 than in model A.8. When SMD is 5, as in model A.8, the standard errors of  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are slightly downward biased and the RB is .95 for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters when  $n \geq 50$ .

Second, the effect of construct on the RB is examined, comparing the RB in model B.8 with the RB in model A.8 (see Table 6.21 before). When SMD is 1, the RB for  $\psi_{00}$  in model B.8 increases from .85 to .96 when the sample size increases from  $n=50$  to  $n=1000$ . For  $\psi_{11}$  parameter, the RB is stable and lower than .85, and for  $\psi_{01}$  it slightly increases from .72 to .76 when the sample size increases from  $n=50$  to  $n=1000$ .

When SMD is 2, the bias of standard error decreases when the sample size increases. As can be seen from Figures 6.39., 6.40 and 6.41, the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 195$ , for  $\psi_{11}$  when  $n \geq 215$ , and for  $\psi_{01}$  when  $n \geq 280$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 115 and 130 larger and 60 smaller, respectively, in model B.8 than in model A.8. The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 420$ , for  $\psi_{11}$  when  $n \geq 470$ , and for  $\psi_{01}$  when  $n \geq 525$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 35, 215 and 45 larger, respectively, in model B.8 than in model A.8.

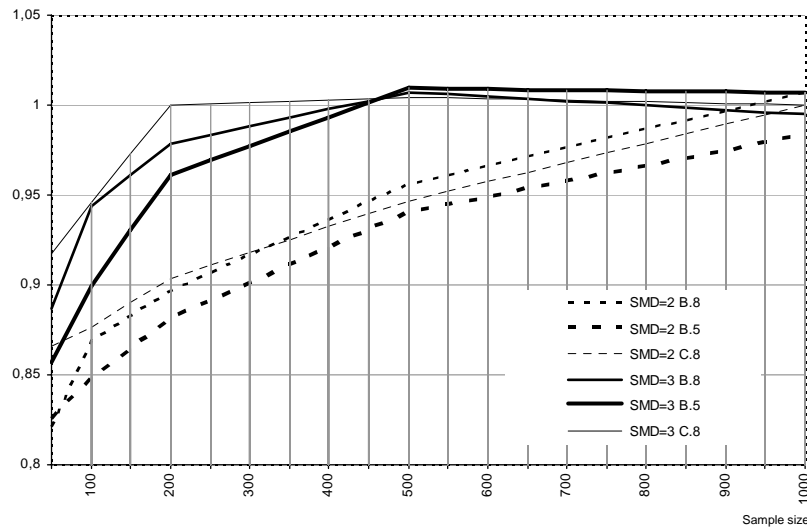


Figure 6.40. The RB for  $\psi_{11}$  parameter in model B.8, B.5 and C.5 when SMD is 2 or 3.

When SMD is 3, the RB is greater than .90 for  $\psi_{11}$  when  $n \geq 60$  and for  $\psi_{01}$  when  $n \geq 80$ . The sample size for  $\psi_{01}$  parameter is 5 larger in model B.8 than in model A.8. For  $\psi_{00}$  parameter, the RB is .90 when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 90$ , for  $\psi_{11}$  when  $n \geq 120$ , and for  $\psi_{01}$  when  $n \geq 125$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 40 smaller, 50 and 25 larger, respectively, in model B.8 than in model A.8.

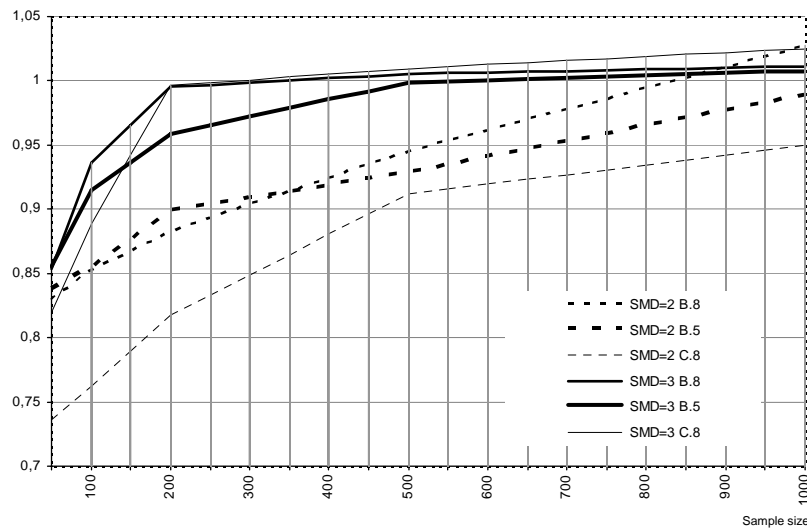


Figure 6.41. The RB for  $\psi_{01}$  parameter in model B.8, B.5 and C.5 when SMD is 2 or 3.

As in model A.8, when SMD is 4, the RBs are high in model B.8, being .94 for  $\psi_{00}$ , .92 for  $\psi_{11}$ , and .91 for  $\psi_{01}$  when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 60$ , for  $\psi_{11}$  when  $n \geq 80$ , and for  $\psi_{01}$  when  $n \geq 80$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 25 lower, 15 and 5 larger, respectively, in model B.8 than in model A.8. As in model A.8, when SMD is 5, the standard errors of  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  in model B.8 are slightly downward biased when  $n \geq 50$ , the RBs for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  in model B.8 being .95.

Third, the effect of construct on the RB is examined, comparing the RB in model B.5 to the RB in model A.5 (see Table 6.21 above). When SMD is 1, the RB for  $\psi_{00}$  in model B.5 increases from .84 to .95 when the sample size increases from  $n=50$  to  $n=1000$ . For  $\psi_{11}$ , the RB is stable and lower than .89, and for  $\psi_{01}$  it

slightly increases from .76 to .80 when the sample size increases from  $n=50$  to  $n=1000$ .

When SMD is 2, the bias of standard error decreases when the sample size increases. As can be seen from Figures 6.39. - 6.41., the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 270$ , for  $\psi_{11}$  when  $n \geq 295$ , and for  $\psi_{01}$  when  $n \geq 210$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 135 and 205 larger and 255 smaller, respectively, in model B.5 than in model A.5. The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 585$ , for  $\psi_{11}$  when  $n \geq 610$ , and for  $\psi_{01}$  when  $n \geq 675$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 185 and 320 larger and 255 smaller, respectively, in model B.5 than in model A.5.

When SMD is 3, the RB is greater than .90 for  $\psi_{00}$  when  $n \geq 70$ , for  $\psi_{11}$  when  $n \geq 100$ , and for  $\psi_{01}$  when  $n \geq 90$ . The sample sizes for  $\psi_{00}$  and  $\psi_{01}$  parameters are 5-10 smaller in model B.5 than in model A.5. The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 135$ , for  $\psi_{11}$  when  $n \geq 180$ , and for  $\psi_{01}$  when  $n \geq 180$ . These sample sizes for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are 30 smaller, 60 and 20 larger, respectively, in model B.5 than in model A.5.

As in model A.5, when SMD is 4, the RBs are high in model B.5, being .92 for  $\psi_{00}$ , .91 for  $\psi_{11}$ , and .90 for  $\psi_{01}$  when  $n=50$ . The RB is greater than .95 for  $\psi_{00}$  when  $n \geq 75$ , for  $\psi_{11}$  when  $n \geq 80$ , and for  $\psi_{01}$  when  $n \geq 90$ . The sample sizes for  $\psi_{00}$  and  $\psi_{01}$  parameters are 20 smaller and 5 larger, respectively, in model B.5 than in model A.5. As in model A.5, when SMD is 5, the standard errors of  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  are slightly downward biased in model B.5, and the RB for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  in model B.5 is .95 when  $n \geq 50$ .

#### **6.3.3.4. Results of RB for $\theta_1$ , $\theta_2$ , $\theta_3$ and $\theta_4$**

The standard errors of  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters (see Table 6.22.) are mainly downward biased in model A.8, but the RB is mainly greater than .95 and always greater than .93. In model A.8, the RB for  $\theta_1$  is lower than .95 when SMD is 1, 2 or 5 and  $n=50$ , and the RB for  $\theta_4$  is lower than .95 when  $n=50$ .

*The effect of reliability on the RB for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$*

When comparing the RBs for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters in model A.5, the results are very similar to those in model A.8. The standard errors of  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  are mainly downward biased also in model A.5, as was the case in model A.8, but in model A.5 the RBs are mainly greater than .95 and always greater than .92. In model A.5, the RB for  $\theta_1$  is lower than .95 when SMD is 1 or 2 and  $n=50$ , and the RB for  $\theta_4$  is lower than .95 when  $n=50$ .

*The effect of additional measurements on the RB for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$*

When comparing the RBs for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  in model A.5\*, the results are very similar to those in model A.5. The standard errors of  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  are mainly downward biased also in model A.5\*, as is the case in model A.5, but in model A.5\* the RBs are mainly greater than .95 and always greater than .93. In model A.5\*, the RB for  $\theta_1$  is lower than .95 when SMD is 1, 2 or 5 and  $n=50$ , and the RB for  $\theta_3$  and  $\theta_4$  is between .94 and .95 when  $n=50$ .

*The effect of model construct on the RB for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$*

When comparing the RBs for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters in model C.8, the results are very similar to those in model A.8. The standard errors of  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  are mainly downward biased in model C.8, as was the case in model A.8, but in model C.8 the RBs are mainly greater than .95 and always greater than .92. In model C.8., the RB for  $\theta_1$  is lower than .95 when SMD is 1, 2 or 3 and  $n=50$ , and the RB for  $\theta_4$  is between .93 and .95 when  $n=50$ .

When comparing the RBs for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters in model B.8, the results are very similar to those in model A.8. The standard errors of  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  are mainly downward biased in model B.8, as was the case in model A.8, but in model B.8, the RB is mainly greater than .95 and always greater than .93. In model B.8., the RB is lower than .95 for  $\theta_1$  when  $n=50$  and for  $\theta_4$  when SMD is 1 or 2 and  $n=50$ .

Table 6.22. The RB for  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model	A.8				A.5				A.5*				B.8				B.5				C.8				
n	SMD	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_3$	$\theta_5$	$\theta_7$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
50	1	.9319	.9529	.9557	.9369	.9307	.9555	.9567	.9293	.9308	.9533	.9482	.9473	.9319	.9529	.9557	.9369	.9307	.9555	.9567	.9293	.919	.9541	.9533	.9338
100	1	.9699	.9727	.9707	.9616	.9693	.9646	.9753	.9638	.9779	.9762	.9637	.9739	.9699	.9727	.9707	.9616	.9693	.9646	.9753	.9638	.9729	.9717	.9728	.9635
200	1	.985	.9891	.987	.9911	.9811	.9852	.9901	.9918	.9953	.9866	.9906	.9859	.985	.9891	.987	.9911	.9811	.9852	.9901	.9918	.985	.9898	.9856	.9912
500	1	.9956	.9965	.9931	.9965	1.000	.9961	.9954	.9932	1.0075	.9933	.9985	.9899	.9956	.9965	.9931	.9965	1.000	.9961	.9954	.9932	1.000	.9946	.9935	.9974
1000	1	1.000	1.000	1.0099	.9967	.9979	.9972	1.0094	.9899	1.0045	.9984	1.000	1.0006	1.000	1.000	1.0099	.9967	.9979	.9972	1.0094	.9899	1.0063	1.000	1.0117	.9914
50	2	.9358	.9573	.9584	.9362	.9254	.9542	.9614	.9342	.9473	.9546	.9454	.9464	.945	.963	.9667	.9434	.9477	.9598	.9627	.931	.9227	.9538	.9538	.9345
100	2	.976	.9736	.9717	.9619	.9733	.9679	.9737	.9634	.988	.9783	.9633	.9699	.9702	.9735	.9779	.968	.9700	.9657	.9755	.964	.9817	.9714	.9742	.9662
200	2	.9871	.9894	.9855	.9896	.9842	.9846	.9893	.9937	.9941	.9881	.9902	.9834	.9857	.9911	.9903	.9931	.9837	.9846	.9945	.9939	.9861	.9868	.9857	.9902
500	2	.9979	.9966	.9931	.9965	1.0029	.9943	.9954	.9929	1.0053	.9933	.997	.989	.9886	.9929	.9978	1.000	.9931	.9941	.9988	.9953	1.0041	.9974	.9919	.9948
1000	2	1.006	1.0048	1.0099	.9967	1.001	.9973	1.0093	.9914	.9926	.9984	.999	.9969	.9968	1.0051	1.0095	.9953	.9946	.9986	1.0089	.991	1.0117	1.0038	1.0093	.9915
50	3	.9595	.9684	.956	.9342	.954	.9624	.9612	.9362	.9609	.9563	.9463	.9466	.9445	.9594	.9681	.9567	.9472	.9562	.9618	.9423	.949	.9646	.9579	.9403
100	3	.9831	.978	.9725	.9621	.9769	.9701	.974	.9629	.9837	.9769	.9627	.9688	.9685	.9718	.9774	.9721	.9702	.9659	.9769	.9742	.9825	.974	.9727	.9647
200	3	.9872	.9895	.9855	.9889	.9859	.9854	.9901	.9921	.988	.9881	.9897	.9817	.9844	.991	.9918	.996	.9836	.9845	.9958	.9949	.9903	.9902	.9856	.9913
500	3	.9938	.9933	.9954	.9965	1.000	.9944	.9961	.9922	1.0011	.9945	.9963	.9882	.9843	.9929	1.0022	1.0021	.9893	.9941	.9994	.9968	1.000	.9948	.9935	.994
1000	3	1.0059	1.0048	1.0099	.9967	1.003	.9987	1.0102	.9919	.9925	.9984	.9979	.9963	.9936	1.000	1.0093	.994	.9913	1.000	1.007	.9882	1.0113	1.0074	1.007	.9915
50	4	.9568	.9705	.9536	.9342	.9617	.9681	.9618	.9361	.9612	.9533	.9451	.9474	.9463	.9592	.9695	.9587	.9488	.9555	.9665	.952	.961	.9672	.9551	.9391
100	4	.975	.9789	.9704	.9595	.9801	.972	.9727	.961	.9822	.9758	.9623	.9685	.9656	.9717	.977	.9714	.9664	.964	.9789	.9789	.9879	.9805	.9733	.9652
200	4	.9827	.9871	.984	.9881	.9842	.9852	.9885	.99	.9863	.9888	.9888	.9819	.9844	.9888	.989	.9945	.9832	.9838	.9935	.9933	.985	.9868	.9856	.9901
500	4	.9873	.9898	.9931	.9965	.9944	.9953	.9974	.9918	1.0022	.9945	.997	.9881	.9844	.9929	1.000	1.0011	.9878	.994	1.0006	.9968	.992	.9921	.9935	.9948
1000	4	1.0061	1.0098	1.0099	.9967	1.004	1.0027	1.0112	.9924	.9923	.9984	.998	.9963	.9936	1.000	1.0126	.9954	.9903	1.000	1.0078	.986	1.0087	1.0037	1.0093	.9927
50	5	.9479	.9653	.952	.9322	.9575	.966	.9556	.931	.9490	.9506	.9448	.9459	.943	.9568	.963	.9558	.9442	.952	.9644	.9555	.955	.9686	.9553	.9395
100	5	.9691	.9769	.9684	.9589	.978	.9728	.9728	.9607	.9801	.9747	.962	.9682	.9673	.9732	.9783	.9669	.9673	.9646	.98	.9772	.978	.9788	.9718	.9634
200	5	.9847	.9912	.9855	.9881	.985	.9872	.9884	.9902	.9873	.9887	.9888	.9822	.9844	.991	.9887	.9929	.9832	.9837	.9899	.9927	.9855	.9916	.9856	.9901
500	5	.9803	.993	.9931	.9965	.9891	.9951	.9974	.9925	1.0045	.9933	.9978	.9877	.9843	.9929	.9978	1.0011	.9871	.994	.9981	.9957	.9874	.9920	.9919	.9939
1000	5	1.000	1.005	1.0099	.9967	1.001	1.0028	1.0122	.9924	.9937	.9953	.9979	.9969	.9968	1.000	1.0096	.9968	.9903	.9986	1.0089	.987	1.0091	1.0076	1.0093	.9914



When comparing the RBs for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  parameters in model B.5, the results are very similar to those in model A.5. The standard errors of  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$  are mainly downward biased in model B.5, as is the case in model A.5, but in model B.5, the RB is mainly greater than .95 and always greater than .929. In model B.5, the RB is lower than .95 for  $\theta_1$  when  $n=50$  and for  $\theta_4$  when SMD is 1, 2 or 3 and  $n=50$ .

#### **6.3.3.5. A summary of the results of RB**

In model A.8, the standard error estimates for mean, variance and covariance parameters are badly downward biased when SMD is 1. In the case SMD is 2, the bias of standard error rapidly decreases when the sample size increases and the RB is, on average, greater than .95 when the sample size is 400-500 for the intercept means, 700-800 for the slope means, 450 for the intercept variance, 250 for the slope variance and 470 for the covariance of intercept and slope. It is noticeable that the RB increases over 1 when the sample size increases to 1000, and is greatest (1.03-1.04) for the mean of intercept. This happens also when SMD is 3, 4 or 5, but in these cases the greatest RB is 1.03.

When SMD is 3, the sample size needed to achieve  $RB=.95$  is 70-130 for parameters of latent components, which can also be seen from Figures 6.42 and 6.43. When SMD is 4 or 5, the critical range for the mean parameters is achieved already with the smallest sample size  $n=50$  and for the variance and covariance parameters when  $n=75$ . For the error variances, the standard errors are slightly downward biased and the RB values are .95, except in the first and the last measurements with  $n=50$  in which cases the RB is at least .93.

When reliability of observed variables decreases (A.5 vs. A.8), the bias of standard error increases. This increase is significant when SMD is 2 and is shown for all mean parameters and covariance parameter. The sample size  $n=1000$  is not enough to achieve  $RB=.95$  in the case of three of the mean parameters and their RB is only .84 - .88. When SMD is 3, the required sample size to achieve  $RB=.95$  is 1.5 – 2.0 times larger for mean parameters in class 2 in model A.5 than in model A.8 as can also be seen from Figure 6.42. The required sample size to achieve  $RB=.95$  in model A.5 for the mean, variance and covariance parameters of latent components is between 130 and 220, whereas in model A.8, the required sample size is 80 – 130. When SMD is 4 and the sample size is  $n=50$ , the RB is on average .02 - .05 smaller in model A.5 than in model A.8. This difference decreases and is nonsignificant when the sample size increases to  $n=100$ . When SMD is 5, the differences in the RB between model A.5 and A.8 are small and nonsignificant. The effect

of reliability on the RB of variances of intercept, slope and error variances are nonsignificant.

The effect of construct on the RB was examined, comparing model C.8 with A.8, B.8 with A.8 and B.5 with A.5. As in model C.8, RB = .95 for mean parameters are not achieved when SMD is 2 in models A.5 and B.5, even when the sample size increases to n=1000. Instead, for the mean of intercept and the variance of slope in the model B.8, the sample size needed to achieve RB=.95 is 1.5 times larger in the model B.8 than in model A.8, whereas for the mean of slope, the variance of intercept and the covariance of intercept and slope, the required sample size is same in both models. When SMD is 3, the sample size needed to achieve RB = .95 in model B.8 or B.5 is 1.1 – 2.3 times larger than in model A.8 or A.5. This proportion is for the mean of intercept 1.3 – 1.9, for the variance of slope 1.6 -2.3 and for the covariance of intercept and slope 1.1 -1.2, whereas for the mean of slope, the required sample size is 0.7 – 1.1, and for the variance of intercept it is 0.7-0.8. When SMD is 4 or 5, differences between models B.8 vs. A.8 or B.5 vs. A.5 in the RB are small (0.01- 0.08). The largest differences are evident when SMD is 4 or 5 and the sample size is 50, in which case the RB in model B.5 is .08 smaller than the RB in model A.5. With all SMDs and with all sample sizes, standard errors of error variances are small.

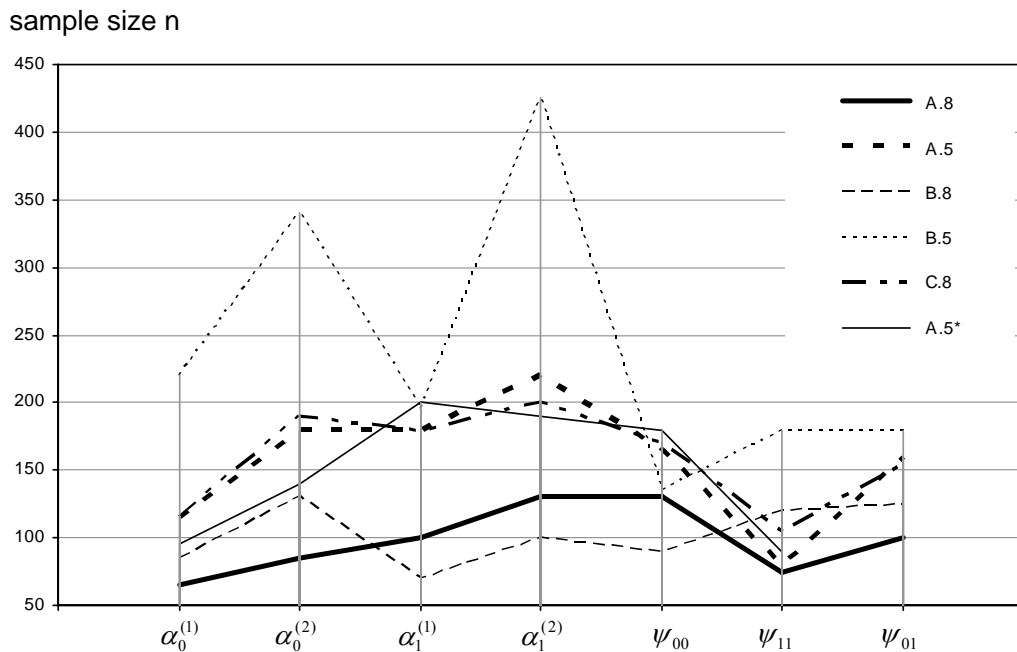


Figure 6.42. The required sample size to achieve RB=.95 when SMD is 3

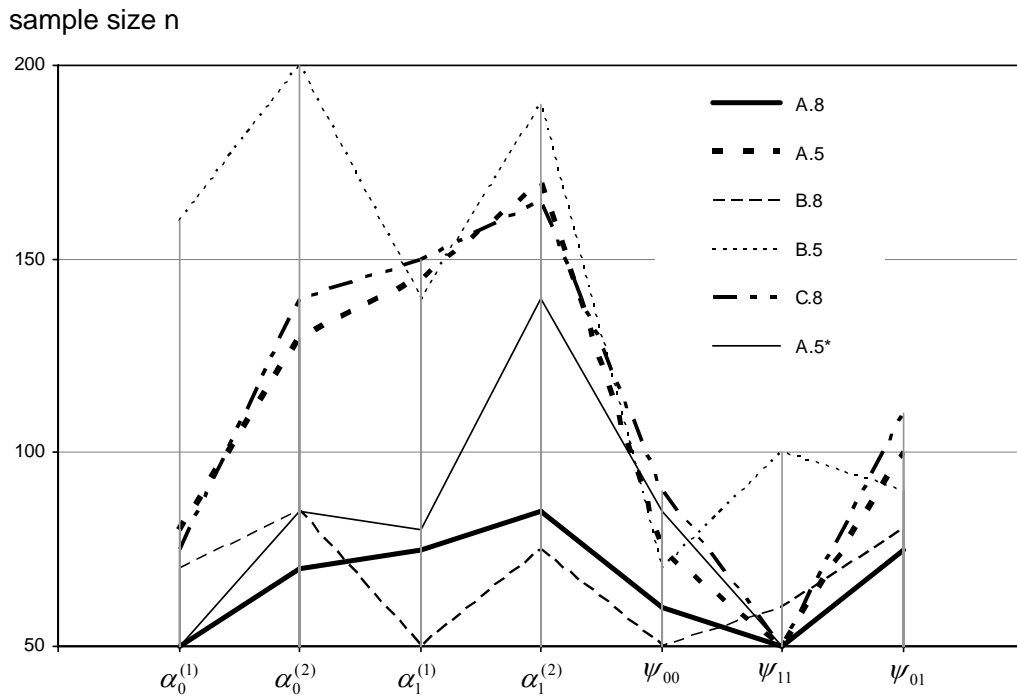


Figure 6.43. The required sample size to achieve RB=.90 when SMD is 3

### 6.3.4. Results of 95 % coverage for parameters

Because the standard error estimates variate around its average value, it is worthwhile to examine the 95 % coverage (referred to later as ‘coverage’) of parameter estimates to evaluate further success of estimation (see section 5.6). Because the standard error was in most cases downward biased, the 95 % coverage is lower than expected .95 value. For unbiased parameter estimation, an expected value of coverage is .937, .922 or .904, with 5, 10 or 15 % downward biased standard error, respectively. Two cut-off values, .90 and .92, are chosen for the 95 % coverage constituting the results. If the 95 % coverage is lower than .92, estimation is seen to be suspicious and if this value is lower than .90, estimation is seen to be poor. In addition to this, linear approximation is produced for sample sizes in which the 95 % coverage is greater than .93. This cut-off value is assumed to be a sign of good estimation.

#### 6.3.4.1. Results of 95 % coverage for $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$

As can be seen from Table 6.23, when SMD is 1, coverages for parameters  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are low and vary between .61 and .79 in all models A.8 – C.8. When SMD is 2, coverage is lower than .90 when the sample size is  $n \leq 200$

in all models A.8 – C.8. When SMD is 3, coverage is lower than .90 when n = 50 in all models except A.8, whereas when the sample size is 500 or 1000, coverage is greater than .93 in all models A.8 – C.8. When SMD is 4 or 5 and n ≥ 100, coverage is greater than .93 in all models A.8 – C.8, except in one case in which coverage is .92.

Table 6.23. 95% coverage for  $\alpha_0^{(1)}, \alpha_0^{(2)}$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8		A.5		A.5*		B.8		B.5		C.8	
n	SMD	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$	$\alpha_0^{(1)}$	$\alpha_0^{(2)}$
50	1	.626	.711	.607	.691	.673	.703	.626	.711	.607	.691	.597	.691
100	1	.646	.726	.623	.696	.677	.675	.646	.726	.623	.696	.609	.703
200	1	.641	.746	.610	.706	.672	.645	.641	.746	.610	.706	.606	.715
500	1	.664	.770	.623	.720	.650	.681	.664	.770	.623	.720	.622	.741
1000	1	.687	.788	.653	.740	.634	.713	.687	.788	.653	.740	.641	.758
50	2	.792	.773	.765	.751	.793	.736	.802	.746	.774	.715	.759	.747
100	2	.833	.832	.793	.793	.807	.798	.818	.789	.783	.743	.786	.795
200	2	.859	.872	.806	.833	.837	.866	.839	.837	.795	.781	.805	.836
500	2	.916	.934	.870	.902	.896	.935	.905	.895	.846	.846	.869	.902
1000	2	.940	.947	.911	.931	.936	.951	.934	.928	.904	.899	.909	.935
50	3	.902	.877	.864	.848	.890	.870	.895	.857	.853	.799	.870	.840
100	3	.932	.923	.907	.898	.922	.924	.927	.902	.887	.849	.907	.895
200	3	.947	.940	.931	.931	.941	.941	.942	.933	.918	.902	.931	.930
500	3	.952	.949	.949	.949	.949	.948	.951	.944	.946	.934	.951	.948
1000	3	.950	.949	.948	.951	.951	.949	.951	.950	.951	.945	.951	.951
50	4	.938	.920	.919	.904	.932	.923	.930	.907	.908	.874	.927	.891
100	4	.947	.940	.943	.934	.940	.933	.944	.932	.935	.918	.944	.936
200	4	.951	.947	.947	.946	.945	.942	.948	.949	.947	.940	.950	.946
500	4	.953	.950	.951	.949	.952	.947	.951	.948	.951	.945	.952	.951
1000	4	.949	.950	.948	.952	.951	.948	.948	.949	.949	.950	.951	.948
50	5	.939	.924	.938	.924	.937	.923	.937	.919	.930	.910	.941	.922
100	5	.949	.939	.950	.938	.942	.937	.945	.936	.943	.932	.947	.942
200	5	.944	.946	.949	.945	.946	.941	.948	.946	.947	.946	.951	.946
500	5	.952	.948	.951	.947	.951	.947	.952	.948	.951	.948	.954	.950
1000	5	.949	.951	.949	.951	.947	.949	.950	.947	.948	.949	.950	.948

Note. Cells highlighted in grey indicate that the 95 % coverage is lower than .92. If the value in the grey-coloured cell is bolded, this means that the value is lower than .90.

*Model A.8 -  $\alpha_0^{(1)}$*

In model A.8, when SMD is 2, coverage for  $\alpha_0^{(1)}$  increases from .792 to .916, when the sample size increases from 50 to 500 and is greater than .93 when

$n \geq 790$ . When SMD is 3, coverage is .902 when  $n = 50$  and is greater than .93 when  $n \geq 95$ . When SMD is 4 or 5, coverage is greater than .93 when  $n \geq 50$ .

*Model A.8 -  $\alpha_0^{(2)}$*

When SMD is 2, coverage for  $\alpha_0^{(2)}$  increases from .773 to .872 when the sample size increases from 50 to 200 and is greater than .93 when  $n \geq 480$ . When SMD is 3, coverage increases from .877 to .923 when the sample size increases from 50 to 100 and is greater than .93 when  $n \geq 140$ . When SMD is 4, coverage is .92 when  $n = 50$  and is greater than .93 when  $n \geq 75$ . When SMD is 5, coverage is .924 when  $n = 50$  and is greater than .93 when  $n \geq 70$ .

*The effect of reliability on the coverage for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

In model A.5 (see Table 6.23), coverage for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters are lower than in model A.8. This is also seen when comparing model B.5 with model B.8.

*Model A.5 vs. model A.8 -  $\alpha_0^{(1)}$*

When SMD is 2, coverage for  $\alpha_0^{(1)}$  in model A.5 increases from .765 to .911 when the sample size increases from 50 to 1000 and is .03 - .05 lower than in model A.8. When SMD is 3, coverage increases from .864 to .907 when the sample size increases from 50 to 100 and is .025 - .038 lower than in model A.8. Coverage is greater than .93 when  $n \geq 195$ , which  $n$  is 100 greater than in model A.8. When SMD is 4, coverage is .919 when  $n = 50$  which coverage is .019 lower than in model A.8. Coverage is greater than .93 when  $n \geq 75$ , which  $n$  is at least 25 greater than in model A.8. As in model A.8, when SMD is 5, coverage in model A.5 is greater than .93 when  $n \geq 50$ .

*Model A.5 vs. model A.8 -  $\alpha_0^{(2)}$*

When SMD is 2, coverage for  $\alpha_0^{(2)}$  in model A.5 increases from .751 to .931 when the sample size increases from 50 to 1000 and is .016 - .039 lower than in model A.8. When SMD is 3, coverage increases from .848 to .931 when the sample size increases from 50 to 200 and is .009 - .029 lower than in model A.8. Coverage is greater than .93 when  $n \geq 195$ , which  $n$  is 55 greater than in model A.8. When SMD is 4, coverage is .904 when  $n = 50$  which coverage is .016 lower than in model A.8. Coverage is greater than .93 when  $n \geq 95$ , which  $n$  is 20 greater than in model A.8. As in model A.8, when SMD is 5, coverage is .924 when  $n = 50$  and is greater than .93 when  $n \geq 70$ .

*The effect of additional measurements on the coverage for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

In model A.5\* (see Table 6.23), coverage for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters are clearly greater than in model A.5.

*Model A.5\* vs. model A.5 -  $\alpha_0^{(1)}$*

When SMD is 2, coverage for  $\alpha_0^{(1)}$  in model A.5\* increases from .793 to .896 when the sample size increases from 50 to 500 and is .028 - .031 greater than in model A.5. When  $n \geq 925$ , coverage in model A.5\* is greater than .93. When SMD is 3, coverage increases from .890 to .922 when the sample size increases from 50 to 100 and is .026 - .015 greater than in model A.5. Coverage is greater than .93 when  $n \geq 140$ , which  $n$  is 55 lower than in model A.5. When SMD is 4, coverage in model A.5\* is greater than .93, with the smallest sample size  $n = 50$ , whereas in model A.5, the required sample size is at least 25 greater. As in model A.8, when SMD is 5, coverage in model A.5 is greater than .93 when  $n \geq 50$ .

*Model A.5\* vs. model A.5 -  $\alpha_0^{(2)}$*

When SMD is 2, coverage for  $\alpha_0^{(2)}$  in model A.5\* increases from .736 to .866 when the sample size increases from 50 to 200. When  $n = 50$  this coverage is .015 lower, and when  $n = 200$  it is .033 greater, than in model A.5. Coverage in model A.5\* is greater than .93 when  $n \geq 480$ , which sample size is over 520 lower than in model A.5. When SMD is 3, coverage increases from .870 to .924 when the sample size increases from 50 to 100 and is .022 - .026 greater than in model A.5. Coverage is greater than .93 when  $n \geq 135$ , which  $n$  is 60 lower than in model A.5. When SMD is 4, coverage is .923 which coverage is .019 greater than in model A.5. Coverage is greater than .93 when  $n \geq 65$ , which  $n$  is 30 lower than in model A.5. When SMD is 5, coverage is .923 which coverage is .001 lower than in model A.5. Coverage is greater than .93 when  $n \geq 75$ , which  $n$  is 5 greater than in model A.5.

*The effect of model construct on the coverage for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$*

When comparing coverages for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  between different models, they are clearly lower in model C.8 than in model A.8. On the contrary, coverages for  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  are slightly lower in model B.8 than in model A.8. This is also seen when comparing coverages between models B.5 and A.5.

*Model C.8 vs. model A.8 -  $\alpha_0^{(1)}$*

When SMD is 2, coverage for  $\alpha_0^{(1)}$  in model C.8 increases from .759 to .909 when the sample size increases from 50 to 1000 and is .031 - .054 lower than

in model A.8. When SMD is 3, coverage increases from .870 to .907 when the sample size increases from 50 to 100 and is .025 - .032 lower than in model A.8. Coverage is greater than .93 when  $n \geq 195$ , which  $n$  is 100 greater than in model A.8. When SMD is 4, coverage is .927 when  $n = 50$ , which coverage is .011 lower than in model A.8, and the coverage is greater than .93 when  $n \geq 60$ , which  $n$  is 10 greater than in model A.8. As in model A.8, when SMD is 5, coverage in model C.8 is greater than .93 when  $n \geq 50$ .

*Model C.8 vs. model A.8 -  $\alpha_0^{(2)}$*

When SMD is 2, coverage for  $\alpha_0^{(2)}$  in model C.8 increases from .747 to .902 when the sample size increases from 50 to 500 and is .026 - .037 lower than in model A.8. Coverage in model C.8 is greater than .93 when  $n \geq 925$ , which  $n$  is 445 greater than in model A.8. When SMD is 3, coverage increases from .840 to .930 when the sample size increases from 50 to 200 and is .010 - .037 lower than in model A.8. Coverage is greater than .93 when  $n = 200$ , which  $n$  is 105 greater than in model A.8. When SMD is 4, coverage is .891 when  $n = 50$ , which coverage is .029 lower than in model A.8. The coverage is greater than .93 when  $n \geq 95$ , which  $n$  is 20 greater than in model A.8. When SMD is 5, coverage is .922 when  $n = 50$ , which coverage is .002 lower than in model A.8. The coverage is greater than .93 when  $n \geq 70$ , which  $n$  is the same than in model A.8.

*Model B.8 vs. model A.8 -  $\alpha_0^{(1)}$*

When SMD is 2, coverage for  $\alpha_0^{(1)}$  in model B.8 increases from .802 to .905 when the sample size increases from 50 to 500 and is .032 - 0.043 lower than in model A.8. Coverage in model B.8 is greater than .93 when  $n \geq 930$ , which  $n$  is 140 greater than in model A.8. When SMD is 3, coverage increases from .895 to .927 when the sample size increases from 50 to 100 and is .020 - .025 lower than in model A.8. Coverage is greater than .93 when  $n \geq 120$ , which  $n$  is 25 greater than in model A.8. As in model A.8, when SMD is 4 or 5, coverage in model B.8 is greater than .93 when  $n \geq 50$ .

*Model B.8 vs. model A.8 -  $\alpha_0^{(2)}$*

When SMD is 2, coverage for  $\alpha_0^{(2)}$  in model B.8 increases from .747 to .902 when the sample size increases from 50 to 500 and is .026 - .037 lower than in model A.8. Coverage is greater than .93 when  $n \geq 925$ , which  $n$  is 445 greater than in model A.8. When SMD is 3, coverage increases from .840 to .930 when the sample size increases from 50 to 200 and is .010 - .037 lower than in model A.8. Coverage is greater than .93 when the sample size is 200 greater than in model A.8. When SMD is 4, coverage is .891 when  $n = 50$ , which coverage is .029 lower than in model A.8. The coverage is greater than .93 when  $n \geq 95$ , which  $n$  is 20 greater than in model A.8. When SMD is 5,

coverage is .922 when  $n = 50$ , which coverage is .002 lower than in model A.8. The coverage is greater than .93 when  $n \geq 70$ , which  $n$  is the same than in model A.8.

*Model B.5 vs. model A.5 -  $\alpha_0^{(1)}$*

When SMD is 2, coverage for  $\alpha_0^{(1)}$  in model B.5 increases from .774 to .904 when the sample size increases from 50 to 1000 and is .009 – .024 lower than in model A.5. When SMD is 3, coverage increases from .853 to .918 when the sample size increases from 50 to 200 and is .011 - .020 lower than in model A.5. Coverage is greater than .93 when  $n \geq 330$ , which  $n$  is 135 greater than in model A.5. When SMD is 4, coverage is .908, which coverage is .011 lower than in model A.5 when  $n = 50$ . Coverage increases to .93 when the sample size increases to  $n \geq 90$ , which  $n$  is 15 greater than in model A.5. As in model A.5, when SMD is 5, coverage in model B.5 is greater than .93 when  $n \geq 50$ .

*Model B.5 vs. model A.5 -  $\alpha_0^{(2)}$*

When SMD is 2, coverage for  $\alpha_0^{(2)}$  in model B.5 increases from .715 to .899 when the sample size increases from 50 to 1000 and is .032 – .056 lower than in model A.5. When SMD is 3, coverage increases from .799 to .902 when the sample size increases from 50 to 200 and is .029 - .049 lower than in model A.5. Coverage is greater than .93 when  $n \geq 460$ , which  $n$  is 265 greater than in model A.5. When SMD is 4, coverage increases from .874 to .918 when the sample size increases from 50 to 100 and is .016 - .030 lower than in model A.5. Coverage is greater than .93 when  $n \geq 155$ , which  $n$  is 60 greater than in model A.5. When SMD is 5, coverage is .910, which coverage is .014 lower than in model A.5 when  $n = 50$ . Coverage increases to .93 when the sample size increases to  $n \geq 95$ , which  $n$  is 15 greater than in model A.5.

**6.3.4.2. Results of 95 % coverage for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$**

As can be seen from Table 6.24, when SMD is 1, coverages for parameters  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  are low and vary between .641 and .874 in all of the models A.8 – C.8. When SMD is 2, coverage is lower than .90 when  $n = 50, 100$  or 200 in all models A.8 – C.8, with one exception. When SMD is 3, coverage is lower than .90 when  $n = 50$  in all models A.8 – C.8 with one exception, whereas when  $n = 500$  or 1000 coverage is greater than .93 in all models A.8 – C.8. When SMD is 4 or 5 and  $n \geq 100$ , coverage is greater than .93 in all models A.8 – C.8.



*Model A.8 -  $\alpha_1^{(1)}$*

In model A.8, when SMD is 2, coverage for  $\alpha_1^{(1)}$  increases from .805 to .910 when the sample size increases from 50 to 500 and is greater than .93 when  $n \geq 720$ . When SMD is 3, coverage is .894 when  $n = 50$  and greater than .93 when  $n \geq 125$ . When SMD is 4, coverage is .924 when  $n = 50$  and greater than .93 when  $n \geq 65$ . When SMD is 5, coverage is greater than .93 when  $n \geq 50$ .

*Model A.8 -  $\alpha_1^{(2)}$*

In model A.8, when SMD is 2, coverage for  $\alpha_1^{(2)}$  increases from .737 to .928 when the sample size increases from 50 to 1000. When SMD is 3, coverage increases from .856 to .905 when the sample size increases from 50 to 100 and is greater than .93 when  $n \geq 140$ . When SMD is 4, coverage is .909 when  $n = 50$  and greater than .93 when  $n \geq 95$ . When SMD is 5, coverage is .920 when  $n = 50$  and greater than .93 when  $n \geq 85$ .

*The effect of reliability on the coverage for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$*

In model A.5 (see Table 6.24), coverages for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters are lower than in model A.8. This effect of reliability is seen also when comparing model B.5 with model B.8.

*Model A.5 vs. model A.8 -  $\alpha_1^{(1)}$*

When SMD is 2, coverage for  $\alpha_1^{(1)}$  in model A.5 increases from .781 to .902 when the sample size increases from 50 to 1000 and is .024 - .045 lower than in model A.8. When SMD is 3, coverage increases from .867 to .926 when the sample size increases from 50 to 200 and is .012 - .027 lower than in model A.8. Coverage is greater than .93 when  $n \geq 215$ , which  $n$  is 90 greater than in model A.8. When SMD is 4, coverage is .917 when  $n = 50$ , which coverage is .007 lower than in model A.8. Coverage is greater than .93 when  $n \geq 80$ , which  $n$  is at least 15 greater than in model A.8. When SMD is 5, coverage is .926 when  $n = 50$ , which coverage is .004 lower than in model A.8. Coverage is greater than .93 when  $n \geq 60$ , which  $n$  is 10 greater than in model A.8.

*Model A.5 vs. model A.8 -  $\alpha_1^{(2)}$*

When SMD is 2, coverage for  $\alpha_1^{(2)}$  in model A.5 increases from .712 to .879, when the sample size increases from 50 to 1000 and is .025 - .066 lower than in model A.8. When SMD is 3, coverage increases from .808 to .909 when the sample size increases from 50 to 200 and is .027 - .049 lower than in model A.8. Coverage is greater than .93 when  $n \geq 265$ , which  $n$  is 140 greater than in model A.8. When SMD is 4, coverage increases from .888 to .925 when the sample size increases from 50 to 100 and is .008 - .021 lower than in model

A.8. Coverage is greater than .93 when  $n \geq 105$ , which  $n$  is 40 greater than in model A.8. When SMD is 5, coverage is .915 when  $n = 50$ , which coverage is .005 lower than in model A.8. Coverage is greater than .93 when  $n \geq 90$ , which  $n$  is 5 greater than in model A.8.

Table 6.24. 95% coverage for  $\alpha_1^{(1)}, \alpha_1^{(2)}$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8		A.5		A.5*		B.8		B.5		C.8	
n	SMD	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$
50	1	.725	.686	.715	.680	.733	.718	.725	.686	.715	.680	.720	.675
100	1	.708	.666	.691	.653	.773	.749	.708	.666	.691	.653	.700	.668
200	1	.704	.653	.681	.641	.809	.791	.704	.653	.681	.641	.693	.655
500	1	.719	.659	.702	.654	.853	.834	.719	.659	.702	.654	.701	.664
1000	1	.737	.665	.721	.668	.874	.856	.737	.665	.721	.668	.715	.659
50	2	.805	.737	.781	.712	.833	.785	.800	.763	.774	.719	.778	.701
100	2	.820	.763	.787	.717	.881	.841	.823	.816	.775	.765	.786	.721
200	2	.850	.799	.805	.741	.915	.896	.849	.858	.794	.799	.804	.742
500	2	.910	.889	.866	.823	.950	.944	.904	.916	.851	.870	.860	.823
1000	2	.937	.928	.902	.879	.951	.956	.932	.938	.901	.913	.904	.883
50	3	.894	.856	.867	.808	.911	.873	.890	.867	.850	.815	.861	.812
100	3	.922	.905	.897	.856	.937	.925	.921	.917	.882	.878	.895	.865
200	3	.938	.936	.926	.909	.943	.946	.940	.942	.917	.916	.925	.908
500	3	.946	.946	.950	.942	.949	.951	.946	.944	.943	.938	.944	.936
1000	3	.947	.947	.950	.945	.948	.951	.945	.943	.941	.941	.952	.946
50	4	.924	.909	.917	.888	.932	.912	.922	.910	.905	.880	.915	.892
100	4	.942	.933	.938	.925	.942	.931	.937	.938	.932	.925	.939	.925
200	4	.945	.941	.947	.942	.944	.944	.945	.943	.946	.940	.943	.940
500	4	.947	.946	.950	.946	.949	.949	.946	.943	.948	.945	.950	.942
1000	4	.950	.945	.952	.944	.949	.950	.947	.943	.949	.941	.952	.945
50	5	.930	.920	.926	.915	.931	.917	.930	.924	.929	.910	.905	.917
100	5	.940	.934	.943	.933	.943	.934	.941	.935	.942	.937	.942	.936
200	5	.946	.941	.946	.943	.944	.942	.944	.944	.946	.944	.945	.942
500	5	.946	.944	.949	.944	.947	.947	.948	.943	.947	.943	.948	.940
1000	5	.949	.946	.950	.945	.949	.948	.948	.943	.951	.943	.951	.944

Note. Cells highlighted in grey indicate that the 95 % coverage is lower than .92. If the value in the grey-coloured cell is bolded, this means that the value is lower than .90.

*The effect of additional measurements on the coverage for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$*

In model A.5\* (see Table 6.24), coverages for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters are clearly greater than in model A.5.

*Model A.5\* vs. model A.5 -  $\alpha_1^{(1)}$*

When SMD is 2, coverage for  $\alpha_1^{(1)}$  in model A.5\* increases from .833 to .915 when the sample size increases from 50 to 200 and is .052 - .110 greater than in model A.5. In model A.5\*, coverage is greater than .93 when  $n \geq 245$ , whereas in model A.5 when  $n = 1000$ , which is not enough to reach this coverage. When SMD is 3, coverage is .911 when  $n = 50$ , which coverage is .044 greater than in model A.5. Coverage is greater than .93 when  $n \geq 85$ , which  $n$  is 130 lower than in model A.5. When SMD is 4 or 5, coverage is greater than .932 or .931, respectively, with the smallest sample size  $n = 50$ . These coverages are .015 or .005 greater, respectively, than in model A.5.

*Model A.5\* vs. model A.5 -  $\alpha_1^{(2)}$*

When SMD is 2, coverage for  $\alpha_1^{(2)}$  in model A.5\* increases from .785 to .896 when the sample size increases from 50 to 200 and is .073 - .155 greater than in model A.5. In model A.5\* coverage is greater than .93 when  $n \geq 270$ , whereas in model A.5  $n = 1000$ , which is not enough to reach that coverage. When SMD is 3, coverage in model A.5\* increases from .873 to .925 when the sample size increases from 50 to 100 and is .065 - .069 greater than in model A.5. In model A.5\*, coverage is greater than .93 when  $n \geq 110$ , which  $n$  is 160 lower than in model A.5. When SMD is 4, coverage is .912 when  $n = 50$ , which coverage is .024 greater than in model A.5. Coverage is greater than .93 when  $n \geq 95$ , which  $n$  is 10 lower than in model A.5. When SMD is 5, coverage is .917 when  $n = 50$ , which coverage is .004 greater than in model A.5. Coverage is greater than .93 when  $n \geq 90$ , which  $n$  is the same as in model A.5.

*The effect of model construct on the coverage for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$*

When comparing coverages for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  between different models, they are clearly lower in model C.8 than in model A.8. Coverage for  $\alpha_1^{(1)}$  is almost equal in model A.8 and in model B.8, whereas coverage for  $\alpha_1^{(2)}$  is greater in model B.8 than in model A.8. This is seen also when comparing coverages in model B.5 with model A.5.

*Model C.8 vs. model A.8 -  $\alpha_1^{(1)}$*

When SMD is 2, coverage for  $\alpha_1^{(1)}$  in model C.8 increases from .778 to .904 when the sample size increases from 50 to 1000 and is .027 - .053 lower than

in model A.8. When SMD is 3, coverage increases from .861 to .925 when the sample size increases from 50 to 200 and is .013 - .033 lower than in model A.8. In model C.8, coverage is greater than .93 when  $n \geq 225$ , which  $n$  is 100 greater than in model A.8. When SMD is 4, coverage is .915 when  $n = 50$ , which coverage is .009 greater than in model A.8. Coverage is greater than .93 when  $n \geq 80$ , which  $n$  is 15 lower than in model A.8. When SMD is 5, coverage is .905 when  $n = 50$ , which coverage is .025 lower than in model A.8. Coverage is greater than .93 when  $n \geq 85$ , which  $n$  is 30 greater than in model A.8.

*Model C.8 vs. model A.8 -  $\alpha_1^{(2)}$*

When SMD is 2, coverage for  $\alpha_1^{(2)}$  in model C.8 increases from .701 to .883 when the sample size increases from 50 to 1000 and is .036 - .066 lower than in model A.8. When SMD is 3, coverage increases from .812 to .908 when the sample size increases from 50 to 200 and is .028 - .044 lower than in model A.5. In model C.8, coverage is greater than .93 when  $n \geq 280$ , which  $n$  is 140 greater than in model A.8. When SMD is 4, coverage is .892 when  $n = 50$ , which coverage is .017 lower than in model A.8. Coverage is greater than .93 when  $n \geq 115$ , which  $n$  is 20 greater than in model A.8. When SMD is 5, coverage is .917 when  $n = 50$ , which coverage is .003 lower than in model A.8. Coverage is greater than .93 when  $n \geq 85$ , which  $n$  is the same as in model A.8.

*Model B.8 vs. model A.8 -  $\alpha_1^{(1)}$*

When SMD is 2, coverage for  $\alpha_1^{(1)}$  in model B.8 increases from .800 to .904 when the sample size increases from 50 to 500 and is .001 - .006 lower than in model A.8. Coverage is greater than .93 when  $n \geq 780$ , which  $n$  is 60 greater than in model A.8. When SMD is 3, coverage increases from .890 to .921 when the sample size increases from 50 to 100 and is .001 - .004 lower than in model A.8. Coverage is greater than .93 when  $n \geq 125$ , which  $n$  is the same as in model A.8. When SMD is 4, coverage is .922 when  $n = 50$ , which coverage is .002 lower than in model A.8. Coverage is greater than .93 when  $n \geq 75$ , which  $n$  is 10 greater than in model A.8. When SMD is 5, coverage is .930 when  $n = 50$ , which coverage is equal to the coverage in model A.8.

*Model B.8 vs. A.8 -  $\alpha_1^{(2)}$*

When SMD is 2, coverage for  $\alpha_1^{(2)}$  in model B.8 increases from .763 to .916 when the sample size increases from 50 to 500 and is .007 - .012 greater than in model A.8. Coverage is greater than .93 when  $n \geq 690$ , which  $n$  is 30 lower than in model A.8. When SMD is 3, coverage increases from .867 to .917 when the sample size increases from 50 to 100 and is .015 - .017 greater than in model A.8. Coverage is greater than .93 when  $n \geq 125$ , which  $n$  is 15

lower than in model A.8. When SMD is 4, coverage is .910 when  $n = 50$ , which coverage is .001 greater than in model A.8. Coverage is greater than .93 when  $n \geq 85$ , which  $n$  is 10 lower than in model A.8. When SMD is 5, coverage is .924 when  $n = 50$ , which coverage is .004 lower than in model A.8. Coverage is greater than .93 when  $n \geq 75$ , which  $n$  is 10 lower than in model A.8.

### 6.3.4.3. Results of 95 % coverage for $\psi_{00}$ , $\psi_{11}$ and $\psi_{01}$

As can be seen from Table 6.25, when SMD is 1, coverages for parameters  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  are low and vary between .641 and .864 in all of the models A.8 – C.8, with one exception concerning  $\psi_{11}$  parameter in model A.5\*. When SMD is 2 and  $n = 50$ , coverage is lower than .90 in all models A.8 – C.8 with one exception. When SMD is 2 and the sample size  $n \geq 100$  or  $n \geq 200$ , coverage in all models is greater than .90 or .92, respectively. When SMD is 4 or 5, coverage is greater than .90 when  $n \geq 100$ , and greater than .92 when  $n \geq 200$ , in all models A.8 – C.8 .

#### *Model A.8 - $\psi_{00}$*

When SMD is 2, coverage for  $\psi_{00}$  in model A.8 increases from .753 to .907 when the sample size increases from 50 to 500 and coverage is greater than .93 when  $n \geq 710$  . When SMD is 3, coverage increases from .843 to .920 when the sample size increases from 50 to 200 and is greater than .93 when  $n \geq 250$  . When SMD is 4, coverage increases from .880 to .913 when the sample size increases from 50 to 100 and is greater than .93 when  $n \geq 145$  . When SMD is 5, coverage increases from .889 to .917 when the sample size increases from 50 to 100 and is greater than .93 when  $n \geq 140$  .

#### *Model A.8 - $\psi_{11}$*

When SMD is 2, coverage for  $\psi_{11}$  in model A.8 increases from .818 to .896 when the sample size increases from 50 to 200 and coverage is greater than .93 when  $n \geq 445$  . When SMD is 3, coverage increases from .879 to .917 when the sample size increases from 50 to 100 and is greater than .93 when  $n \geq 135$  . When SMD is 4, coverage increases from .896 to .921 when the sample size increases from 50 to 100 and is greater than .93 when  $n \geq 125$  . When SMD is 5, coverage increases from .899 to .921 when the sample size increases from 50 to 100 and is greater than .93 when  $n \geq 130$  .

#### *Model A.8 - $\psi_{01}$*

When SMD is 2, coverage for  $\psi_{01}$  in model A.8 increases from .781 to .924 when the sample size increases from 50 to 1000. When SMD is 3, coverage increases from .847 to .896 when the sample size increases from 50 to 100 and is greater than .93 when  $n \geq 150$ . When SMD is 4, coverage is .903 when  $n = 50$  and is greater than .93 when  $n \geq 90$ . When SMD is 5, coverage is .926 when  $n = 50$  and is greater than .93 when  $n \geq 65$ .

#### *The effect of reliability on the coverage for $\psi_{00}$ , $\psi_{11}$ and $\psi_{01}$*

Coverages for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters are, on average, slightly lower in model A.5 than in model A.8, as can be seen from the grey-coloured cells in Table 6.25. This effect of reliability is seen also when comparing model B.5 with model B.8. The results are not consistent suggesting that results depend on the parameter and SMD.

#### *Model A.5 vs. model A.8 - $\psi_{00}$*

When SMD is 2, coverage for  $\psi_{00}$  in model A.5 increases from .759 to .901 when the sample size increases from 50 to 1000 and is .004 - .039 lower than in model A.8. When SMD is 3, coverage increases from .841 to .916 when the sample size increases from 50 to 200 and is .002 - .013 lower than in model A.8. Coverage is greater than .93 when  $n \geq 265$ , which  $n$  is 10 greater than in model A.8. When SMD is 4, coverage increases from .884 to .916 when the sample size increases from 50 to 100 and is unexceptionally .003 - .004 greater than in model A.8. Coverage is greater than .93 when  $n \geq 135$ , which  $n$  is 10 lower than in model A.8. When SMD is 5, coverage increases from .901 to .925 when the sample size increases from 50 to 100 and is also unexceptionally .008 - .012 greater than in model A.8. Coverage is greater than .93 when  $n \geq 120$ , which  $n$  is 20 lower than in model A.8.

#### *Model A.5 vs. A.8 - $\psi_{11}$*

When SMD is 2, coverage for  $\psi_{11}$  in model A.5 increases from .841 to .928 when the sample size increases from 50 to 500. When  $n = 50$ , coverage in model A.5 is .023 greater than in model A.8 and, when the sample size increase to  $n = 200$ , coverage changes to be .010 lower in model A.5 than in model A.8. Coverage in model A.5 is greater than .93 when  $n \geq 590$ , which  $n$  is 145 greater than in model A.8. When SMD is 3, coverage increases from .886 to .919 when the sample size increases from 50 to 100 and is .002 - .007 greater than in model A.8. Coverage is greater than .93 when  $n \geq 125$ , which  $n$  is 10 lower than in model A.8. When SMD is 4, coverage is .913 when  $n = 50$ , which coverage is .017 greater than in model A.8. Coverage is greater than .93 when  $n \geq 90$ , which  $n$  is even 35 lower than in model A.8. When

SMD is 5, coverage is .921 when  $n = 50$ , which coverage is .022 greater than in model A.8. Coverage is greater than .93 when  $n \geq 80$ , which  $n$  is even 50 lower than in model A.8.

*Model A.5 vs. model A.8 -  $\psi_{01}$*

When SMD is 2, coverage for  $\psi_{01}$  in model A.5 increases from .814 to .892 when the sample size increases from 50 to 1000. When  $n = 50$ , coverage in model A.5 is .033 greater than in model A.8, and when the sample size increases to 1000, coverage changes to be .032 lower than in model A.8. When SMD is 3, coverage increases from .848 to .914 when the sample size increases from 50 to 200. When  $n = 50$ , coverage in model A.5 is .001 greater than in model A.8 and, when the sample size increases to 200, coverage changes to be .017 lower than in model A.8. Coverage is greater than .93 when  $n \geq 255$ , which  $n$  is 105 greater than in model A.8. When SMD is 4, coverage increases from .890 to .923 when the sample size increases from 50 to 100 and is .012- .013 lower than in model A.8. Coverage is greater than .93 when  $n \geq 120$ , which  $n$  is 30 greater than in model A.8. When SMD is 5, coverage is .915 when  $n = 50$ , which coverage is .011 lower than in model A.8. Coverage is greater than .93 when  $n \geq 90$ , which  $n$  is 25 greater than in model A.8.

*The effect of additional measurements on the coverage for  $\psi_{00}$  and  $\psi_{11}$*

Coverages for  $\psi_{00}$  and  $\psi_{11}$  parameters are slightly lower in model A.5\* than in model A.5, as can be seen from the grey-coloured cells in Table 6.25. The results are not consistent, suggesting that results depend on SMD.

*Model A.5\* vs. model A5 -  $\psi_{00}$*

When SMD is 2, coverage for  $\psi_{00}$  in model A.5\* increases from .729 to .923 when the sample size increases from 50 to 1000. When  $n = 50$ , coverage in model A.5\* is .030 lower than in model A.5 and, when  $n = 1000$ , coverage changes to be .022 greater than in model A.5. When SMD is 3, coverage increases from .829 to .920 when the sample size increases from 50 to 200. When  $n = 50$ , coverage in model A.5\* is .012 lower than in model A.5, and when  $n = 200$ , coverage changes to be .004 greater than in model A.5. Coverage is greater than .93 when  $n \geq 255$ , which  $n$  is 10 lower than in model A.5. When SMD is 4, coverage increases from .883 to .915 when the sample size increases from 50 to 100 and is .001 lower than in model A.5. Coverage is greater than .93 when  $n \geq 145$ , which  $n$  is 10 greater than in model A.5. When SMD is 5, coverage increases from .894 to .921 when the sample size increases from 50 to 100 and is .004 - .007 lower than in model A.5. Coverage is greater than .93 when  $n \geq 130$ , which  $n$  is 10 greater than in model A.5.

Table 6.25. 95% coverage for  $\psi_{00}, \psi_{11}, \psi_{01}$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model		A.8			A.5			A.5*		B.8			B.5			C.8		
n	SMD	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$	$\psi_{00}$	$\psi_{11}$	$\psi_{01}$
50	1	.545	.775	.765	.614	.814	.805	.570	.716	.545	.775	.765	.614	.814	.805	.515	.784	.762
100	1	.644	.801	.750	.666	.823	.803	.658	.787	.644	.801	.750	.666	.823	.803	.611	.800	.752
200	1	.715	.817	.743	.721	.838	.798	.731	.839	.715	.817	.743	.721	.838	.798	.687	.817	.755
500	1	.774	.841	.749	.771	.849	.801	.755	.882	.774	.841	.749	.771	.849	.801	.753	.837	.747
1000	1	.782	.861	.753	.788	.864	.799	.723	.904	.782	.861	.753	.788	.864	.799	.773	.844	.757
50	2	.753	.818	.781	.759	.841	.814	.729	.792	.771	.752	.859	.793	.785	.862	.724	.810	.774
100	2	.810	.864	.788	.814	.870	.829	.775	.867	.824	.806	.869	.829	.814	.875	.793	.842	.778
200	2	.839	.896	.816	.828	.895	.830	.813	.912	.863	.827	.895	.864	.838	.891	.816	.876	.794
500	2	.907	.938	.888	.871	.928	.859	.882	.943	.899	.891	.932	.888	.865	.919	.866	.922	.840
1000	2	.940	.948	.924	.901	.946	.892	.923	.951	.931	.925	.948	.916	.897	.943	.905	.942	.879
50	3	.843	.879	.847	.841	.886	.848	.829	.865	.851	.838	.887	.849	.832	.875	.865	.865	.831
100	3	.890	.917	.896	.877	.919	.879	.881	.915	.894	.884	.922	.891	.860	.906	.869	.907	.862
200	3	.920	.936	.931	.916	.941	.914	.920	.939	.928	.920	.943	.924	.901	.931	.906	.932	.907
500	3	.939	.946	.946	.938	.947	.942	.939	.939	.942	.942	.949	.942	.935	.946	.935	.947	.937
1000	3	.942	.947	.951	.944	.949	.946	.942	.948	.943	.943	.951	.947	.943	.951	.941	.946	.950
50	4	.880	.896	.903	.884	.913	.890	.883	.888	.885	.885	.914	.888	.875	.897	.868	.896	.890
100	4	.913	.921	.935	.916	.935	.923	.915	.919	.917	.916	.941	.919	.915	.930	.908	.924	.928
200	4	.931	.938	.943	.936	.943	.942	.932	.938	.933	.936	.947	.936	.938	.945	.929	.938	.941
500	4	.941	.947	.950	.942	.949	.947	.946	.940	.941	.948	.951	.944	.948	.948	.940	.947	.946
1000	4	.944	.950	.950	.948	.949	.948	.944	.948	.943	.946	.952	.948	.947	.948	.944	.948	.951
50	5	.889	.899	.926	.901	.921	.915	.894	.891	.888	.900	.931	.900	.902	.911	.890	.905	.916
100	5	.917	.921	.941	.925	.935	.934	.921	.921	.917	.925	.943	.922	.930	.936	.916	.927	.943
200	5	.934	.937	.944	.939	.944	.942	.935	.937	.933	.939	.948	.937	.941	.942	.933	.939	.945
500	5	.939	.947	.948	.941	.948	.947	.948	.941	.939	.948	.948	.943	.948	.947	.940	.947	.947
1000	5	.942	.948	.950	.946	.951	.949	.948	.947	.941	.950	.950	.949	.951	.947	.942	.948	.953

Note. Cells highlighted in grey indicate that the 95 % coverage is lower than .92. If the value in the grey-coloured cell is bolded, this means that the value is lower than .90.

*Model A.5\* vs. A5 -  $\psi_{11}$*

When SMD is 2, coverage for  $\psi_{11}$  in model A.5\* increases from .792 to .912 when the sample size increases from 50 to 200. When  $n = 50$ , coverage in model A.5\* is .049 lower than in model A.5 and, when  $n = 200$ , coverage changes to be .017 greater than in model A.5. Coverage is greater than .93 when  $n \geq 375$ , which  $n$  is 215 lower than in model A.5. When SMD is 3, coverage increases from .865 to .915 when the sample size increases from 50 to 100 and is .004 - .021 lower than in model A.5. Coverage is greater than



.93 when  $n \geq 130$ , which  $n$  is 5 greater than in model A.5. When SMD is 4, coverage increases from .888 to .919 when the sample size increases from 50 to 100 and is .016 - .025 lower than in model A.5. Coverage is greater than .93 when  $n \geq 130$ , which  $n$  is 40 greater than in model A.5. When SMD is 5, coverage increases from .891 to .921 when the sample size increases from 50 to 100 and is .014 - .030 lower than in model A.5. Coverage is greater than .93 when  $n \geq 130$ , which  $n$  is 50 greater than in model A.5.

*The effect of model construct on the coverage for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$*

When comparing coverages for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  between different models, they are mainly lower in model C.8 than in model A.8. Also, the effect of construct is evident when comparing coverages between models B.8 vs. A.8 and between models B.5 vs. A.5.

*Model C.8 vs. model A.8 -  $\psi_{00}$*

When SMD is 2, coverage for  $\psi_{00}$  in model C.8 increases from .724 to .905 when the sample size increases from 50 to 1000 and is .017 - .041 lower than in model A.8. When SMD is 3, coverage increases from .865 to .906 when the sample size increases from 50 to 200. When  $n = 50$ , coverage in model C.8 is .022 greater than in model A.8, and when  $n = 200$ , coverage changes to be .014 lower than in model A.8. Coverage is greater than .93 when  $n \geq 285$ , which  $n$  is 30 greater than in model A.8. When SMD is 4, coverage increases from .868 to .929 when the sample size increases from 50 to 200 and is .001 - .012 lower than in model A.8. Coverage is greater than .93 when  $n \geq 210$ , which  $n$  is 65 greater than in model A.8. When SMD is 5, coverage increases from .890 to .916 when the sample size increases from 50 to 100 and is .001 greater when  $n = 50$ , and .001 lower when  $n = 100$ , than in model A.8. Coverage is greater than .93 when  $n \geq 140$ , which  $n$  is the same as in model A.8.

*Model C.8 vs. model A.8 -  $\psi_{11}$*

When SMD is 2, coverage for  $\psi_{11}$  in model C.8 increases from .810 to .922 when the sample size increases from 50 to 500 and is .008 - .022 lower than in model A.8. Coverage is greater than .93 when  $n \geq 620$ , which  $n$  is 175 greater than in model A.8. When SMD is 3, coverage increases from .865 to .907 when the sample size increases from 50 to 100 and is .010 - .014 lower than in model A.8. Coverage is greater than .93 when  $n \geq 145$ , which  $n$  is 10 greater than in model A.8. When SMD is 4, coverage increases from .896 to .924 when the sample size increases from 50 to 100 and is 0 - .003 greater than in model A.8. Coverage is greater than .93 when  $n \geq 120$ , which  $n$  is 5 lower than in model A.8. When SMD is 5, coverage is .905 when the sample

size is 50 and is .006 greater than in model A.8. Coverage is greater than .93 when  $n \geq 115$ , which  $n$  is 15 lower than in model A.8.

*Model C.8 vs. model A.8 -  $\psi_{01}$*

When SMD is 2, coverage for  $\psi_{01}$  in model C.8 increases from .774 to .879 when the sample size increases from 50 to 1000 and is .007 - .048 lower than in model A.8. When SMD is 3, coverage increases from .831 to .907 when the sample size increases from 50 to 200 and is .016 - .034 lower than in model A.8. Coverage is greater than .93 when  $n \geq 275$ , which  $n$  is 125 greater than in model A.8. When SMD is 4, coverage increases from .890 to .928 when the sample size increases from 50 to 100 and is .007 - .013 lower than in model A.8. Coverage is greater than .93 when  $n \geq 110$ , which  $n$  is 20 greater than in model A.8. When SMD is 5, coverage is .916 when  $n = 50$  and is .010 lower than in model A.8. Coverage is greater than .93 when  $n \geq 75$ , which  $n$  is 10 greater than in model A.8.

*Model B.8 vs. model A.8 -  $\psi_{00}$*

When SMD is 2, coverage for  $\psi_{00}$  in model B.8 increases from .771 to .899 when the sample size increases from 50 to 500. When  $n = 50, 100$  or  $200$ , coverage is .018, .014 or .024 greater, respectively, and when  $n = 500$  it is .008 lower, than in model A.8. Coverage is greater than .93 when  $n \geq 790$ , which  $n$  is 80 greater than in model A.8. When SMD is 3, coverage increases from .851 to .928 when the sample size increases from 50 to 200 and is .004 - .008 greater than in model A.8. Coverage is greater than .93 when  $n \geq 215$ , which  $n$  is 40 lower than in model A.8. When SMD is 4, coverage increases from .885 to .917 when the sample size increases from 50 to 100 and is .004 - .005 greater than in model A.8. Coverage is greater than .93 when  $n \geq 140$ , which  $n$  is 5 lower than in model A.8. When SMD is 5, coverage increases from .888 to .917 when the sample size increases from 50 to 100 and is 0 - .001 lower than in model A.8. Coverage is greater than .93 when  $n \geq 140$ , which  $n$  is the same as in model A.8.

*Model B.8 vs. model A.8 -  $\psi_{11}$*

When SMD is 2, coverage for  $\psi_{11}$  in model B.8 increases from .752 to .925 when the sample size increases from 50 to 1000 and is .024 - .081 lower than in model A.8. When SMD is 3, coverage increases from .838 to .920 when the sample size increases from 50 to 200 and is .012 - .040 lower than in model A.8. Coverage is greater than .93 when  $n \geq 250$ , which  $n$  is 115 greater than in model A.8. When SMD is 4, coverage increases from .885 to .916 when the sample size increases from 50 to 100 and is .006 - .011 lower than in model A.8. Coverage is greater than .93 when  $n \geq 135$ , which  $n$  is 10 greater than in model A.8. When SMD is 5, coverage increases from .900 to .925 when the

sample size increases from 50 to 100 and is .002 - .005 greater than in model A.8. Coverage is greater than .93 when  $n \geq 120$ , which  $n$  is 10 lower than in model A.8.

*Model B.8 vs. model A.8 -  $\psi_{01}$*

When SMD is 2, coverage for  $\psi_{01}$  in model B.8 increases from .859 to .895 when the sample size increases from 50 to 200 and is .078 - .081 greater than in model A.8. In model B.8, coverage is greater than .93 when  $n \geq 485$ , whereas in model A.8, coverage is lower than .93 with the highest sample size  $n = 1000$ . When SMD is 3, coverage in model B.8 increases from .887 to .922 when the sample size increases from 50 to 100 and is .026 - .040 greater than in model A.8. Coverage is greater than .93 when  $n \geq 120$  and is 30 lower than in model A.8. When SMD is 4, coverage is .914 when  $n = 50$ , which coverage is .011 greater than in model A.8. Coverage is greater than .93 when  $n \geq 100$ , which  $n$  is 10 greater than in model A.8. When SMD is 5, coverage is .931 with the smallest sample size  $n = 50$ , which coverage is .005 greater than in model A.8.

*Model B.5 vs. model A.5 -  $\psi_{00}$*

When SMD is 2, coverage for  $\psi_{00}$  in model B.5 increases from .793 to .916 when the sample size increases from 50 to 1000 and is .015 - .036 greater than in model A.5. When SMD is 3, coverage increases from .849 to .924 when the sample size increases from 50 to 200 and is .008 - .014 greater than in model A.5. Coverage is greater than .93 when  $n \geq 235$ , which  $n$  is 30 smaller than in model A.5. When SMD is 4, coverage increases from .888 to .919 when the sample size increases from 50 to 100 and is .003 - .004 greater than in model A.5. Coverage is greater than .93 when  $n \geq 130$ , which  $n$  is 5 lower than in model A.5. When SMD is 5, coverage increases from .900 to .922 when the sample size increases from 50 to 100 and is .001 - .003 lower than in model A.5. Coverage is greater than .93 when  $n \geq 125$ , which  $n$  is 5 greater than in model A.5.

*Model B.5 vs. model A.5 -  $\psi_{11}$*

When SMD is 2, coverage for  $\psi_{11}$  in model B.5 increases from .785 to .897 when the sample size increases from 50 to 1000 and is .049 - .063 lower than in model A.5. When SMD is 3, coverage increases from .832 to .901 when the sample size increases from 50 to 200 and is .040 - .059 lower than in model A.5. Coverage is greater than .93 when  $n \geq 285$ , which  $n$  is 160 greater than in model A.5. When SMD is 4, coverage increases from .875 to .915 when the sample size increases from 50 to 100 and is .020 - .038 lower than in model A.5. Coverage is greater than .93 when  $n \geq 135$ , which  $n$  is 45 greater than in model A.5. When SMD is 5, coverage increases from .902 to .930 when the sample size increases from 50 to 100 and is .005 - .019 lower than in model

A.5. Coverage is greater than .93 when  $n \geq 100$ , which  $n$  is 20 greater than in model A.5.

*Model B.5 vs. model A.5 -  $\psi_{01}$*

When SMD is 2, coverage in model B.5 for  $\psi_{01}$  increases from .862 to .919 when the sample size increases from 50 to 500 and is .046 - .061 greater than in model A.5. In model B.5, coverage is greater than .93 when  $n \geq 730$ , whereas in model A.5, coverage is lower than .93 with the highest sample size  $n = 1000$ . When SMD is 3, coverage in model B.5 increases from .875 to .906 when the sample size increases from 50 to 100 and is .027 greater than in model A.5. Coverage is greater than .93 when  $n \geq 150$  and is 105 lower than in model A.5. When SMD is 4, coverage increases from .897 to .930 when the sample size increases from 50 to 100 and is .007 greater than in model A.5. Coverage is greater than .93 when  $n \geq 100$ , which  $n$  is 20 lower than in model A.5. When SMD is 5, coverage is .911 when  $n = 50$ , which coverage is .004 lower than in model A.5. Coverage is greater than .93 when  $n \geq 90$ , which  $n$  is the same as in model A.5.

**6.3.4.4. Results of 95 % coverage for  $\theta_1, \theta_2, \theta_3,$  and  $\theta_4$**

As can be seen from Table 6.26, coverages for parameters  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  are over .92 in all of the models A.8 – C.8 when  $SMD \geq 1$  and  $n \geq 100$ . The few cases in which coverage is lower than .90 are seen when SMD is 1 and  $n = 50$ .

*A.8 -  $\theta_1, \theta_2, \theta_3,$  and  $\theta_4$*

Coverage is, approximately, greater than .93 for  $\theta_1$  when the sample size is 75-95, for  $\theta_2$  when the sample size is 85-100, for  $\theta_3$  when the sample size is 100 – 105, and for  $\theta_4$  when the sample size is 110 – 115. The trend of these sample sizes depends on SMD: the sample sizes decrease when SMD increases.

*The effect of reliability on the coverage for  $\theta_1, \theta_2, \theta_3,$  and  $\theta_4$*

*Model A.5 vs. model A.8 -  $\theta_1, \theta_2, \theta_3,$  and  $\theta_4$*

Coverage is, approximately, greater than .93 for  $\theta_1$  when the sample size is 95-120, for  $\theta_2$  when the sample size is 95-105, for  $\theta_3$  when the sample size is 110, and for  $\theta_4$  when the sample size is 100 – 115. The trend of these sample sizes depends on SMD: the sample sizes decrease when SMD increases. These sample sizes in model A.8 and A.5 are almost equal. The greatest

differences in coverages are for  $\theta_3$  and  $\theta_4$  parameters. Coverages in model A.5 are greater than .93 when the sample sizes are 5 – 10 greater than in model A.8.

*The effect of additional measurements on the coverage for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , and  $\theta_4$*

*Model A.5\* ( $\theta_1$ ,  $\theta_3$ ,  $\theta_5$ , and  $\theta_7$ ) vs. model A.5 -  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , and  $\theta_4$*

Coverage is, approximately, greater than .93 for  $\theta_1$  when the sample size is 95-120, for  $\theta_2$  when the sample size is 95-105, for  $\theta_3$  when the sample size is 110, and for  $\theta_4$  when the sample size is 100 – 115. These sample sizes in model A.5\* are 0 - 20 greater for  $\theta_1$  in model A.5\* than in model A.5. For  $\theta_2$  the sample size is 5 smaller when SMD is 1, and 0 - 20 greater when SMD is 2 in model A.5\* than in model A.5. For  $\theta_3$ , the sample size to achieve .93 coverage is equal or 5 smaller in model A.5\* than in model A.5. For  $\theta_4$ , the sample size is 0 -15 lower than in model A.5.

*The effect of model construct on the coverage for  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , and  $\theta_4$*

*Model C.8 vs. model A.8 -  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , and  $\theta_4$*

Coverage is, approximately, greater than .93 for  $\theta_1$  when the sample size is 70-100, for  $\theta_2$  when the sample size is 90-95, for  $\theta_3$  when the sample size is 95-110, and for  $\theta_4$  when the sample size is 105 – 110. These sample sizes are smaller or greater in model C.8 than in model A.8, and the greatest differences are lower than 5.

*Model B.8 vs. model A.8 -  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , and  $\theta_4$*

Coverage is, approximately, greater than .93 for  $\theta_1$  when the sample size is 70-95, for  $\theta_2$  when the sample size is 90-100, for  $\theta_3$  when the sample size is 90-105, and for  $\theta_4$  when the sample size is 95 – 110. These sample sizes in model B.8 are for  $\theta_1$  0-5 lower, for  $\theta_2$  5-10 greater, and for  $\theta_3$  and  $\theta_4$  5-10 lower than in model A.8.

Table 6.26. 95% coverage for  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$  parameters in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model	A.8				A.5				A.5*				B.8				B.5				C.8				
n	SMD	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_3$	$\theta_5$	$\theta_7$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
50	1	.906	.913	.917	.913	.904	.911	.911	.907	<b>.896</b>	.915	.913	.905	.906	.913	.917	.913	.904	.911	.911	.907	.902	.913	.915	.910
100	1	.932	.930	.929	.926	.930	.929	.927	.925	.926	.930	.928	.930	.932	.930	.929	.926	.930	.929	.927	.925	.931	.933	.928	.927
200	1	.939	.942	.938	.943	.942	.938	.938	.939	.935	.941	.941	.942	.939	.942	.938	.943	.942	.938	.938	.939	.942	.942	.938	.941
500	1	.945	.946	.948	.949	.948	.945	.947	.946	.947	.948	.946	.947	.945	.946	.948	.949	.948	.945	.947	.946	.945	.944	.946	.948
1000	1	.948	.947	.950	.947	.947	.949	.952	.947	.944	.948	.949	.948	.948	.947	.950	.947	.947	.949	.952	.947	.949	.946	.953	.947
50	2	.909	.918	.919	.917	.908	.910	.913	.910	.904	.915	.909	.910	.916	.920	.915	.914	.918	.918	.913	.905	<b>.899</b>	.915	.918	.913
100	2	.936	.932	.930	.926	.936	.932	.927	.925	.931	.929	.928	.929	.935	.931	.931	.927	.933	.930	.930	.925	.935	.931	.930	.929
200	2	.942	.945	.937	.942	.941	.939	.938	.940	.939	.942	.940	.941	.942	.943	.940	.944	.943	.939	.939	.940	.942	.941	.938	.939
500	2	.946	.947	.948	.949	.948	.946	.947	.945	.949	.944	.946	.948	.945	.948	.947	.947	.946	.945	.945	.947	.948	.946	.947	.946
1000	2	.950	.949	.950	.948	.949	.948	.952	.946	.946	.949	.950	.948	.947	.949	.953	.945	.946	.949	.953	.943	.952	.948	.952	.946
50	3	.918	.923	.918	.917	.918	.915	.916	.914	.910	.915	.910	.910	.921	.919	.918	.914	.922	.917	.910	.908	.911	.919	.921	.915
100	3	.938	.933	.930	.928	.936	.933	.928	.926	.932	.930	.927	.928	.936	.932	.933	.931	.935	.931	.933	.928	.937	.933	.931	.928
200	3	.941	.942	.938	.943	.942	.939	.940	.939	.938	.940	.941	.939	.941	.942	.941	.944	.943	.940	.939	.941	.943	.941	.938	.941
500	3	.947	.946	.948	.949	.948	.944	.947	.946	.945	.945	.946	.947	.945	.947	.948	.947	.947	.945	.947	.946	.948	.945	.947	.945
1000	3	.951	.949	.950	.948	.949	.948	.951	.945	.947	.949	.949	.948	.947	.949	.951	.944	.947	.947	.951	.943	.953	.949	.952	.946
50	4	.924	.922	.918	.921	.924	.919	.917	.914	.913	.917	.910	.910	.925	.919	.918	.918	.923	.918	.912	.914	.918	.920	.918	.917
100	4	.937	.933	.929	.927	.936	.932	.928	.926	.932	.931	.927	.929	.938	.931	.931	.932	.936	.931	.933	.932	.939	.932	.930	.928
200	4	.944	.941	.938	.945	.945	.940	.940	.939	.940	.940	.940	.939	.942	.942	.940	.944	.943	.938	.939	.940	.944	.940	.939	.940
500	4	.945	.946	.948	.949	.947	.945	.946	.945	.945	.947	.946	.947	.945	.946	.949	.946	.947	.945	.947	.945	.945	.946	.946	.946
1000	4	.950	.948	.950	.947	.949	.949	.950	.945	.948	.948	.948	.948	.948	.949	.951	.945	.947	.948	.951	.943	.952	.947	.953	.948
50	5	.924	.919	.918	.918	.926	.918	.916	.914	.913	.915	.910	.910	.925	.919	.918	.920	.922	.917	.912	.917	.925	.921	.916	.919
100	5	.936	.934	.929	.927	.936	.933	.928	.927	.932	.931	.927	.927	.938	.932	.931	.930	.936	.931	.932	.930	.938	.932	.929	.928
200	5	.944	.941	.938	.945	.942	.940	.939	.940	.938	.941	.940	.939	.944	.942	.938	.943	.943	.939	.940	.941	.946	.940	.938	.941
500	5	.944	.947	.947	.949	.945	.946	.947	.945	.947	.947	.946	.947	.945	.946	.948	.946	.947	.947	.946	.945	.944	.947	.947	.946
1000	5	.950	.950	.950	.948	.948	.949	.952	.947	.947	.946	.948	.948	.947	.950	.952	.946	.947	.948	.953	.945	.950	.948	.952	.948

Note. Cells highlighted in grey indicate that the 95 % coverage is lower than .92. If the value in the grey-coloured cell is bolded, this means that the value is lower than .90.

*Model B.5 vs. model A.5 -  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , and  $\theta_4$*

Coverage is, approximately, greater than .93 for  $\theta_1$  when the sample size is 75-100, for  $\theta_2$  when the sample size is 95-105, for  $\theta_3$  when the sample size is 95-115, and for  $\theta_4$  when the sample size is 95 – 120. These sample sizes in model B.5 are for  $\theta_1$  0 - 5 greater, for  $\theta_2$  5 greater, for  $\theta_3$  5 lower, and for  $\theta_4$  0-10 lower than in model A.5.

#### **6.3.4.5 Summary of results of 95 % coverage**

When estimation, in terms of the 95 % coverage, is poor (95 % coverage lower than .90) or suspicious (95 % coverage lower than .92), it is seen most clearly for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters and almost as poor or suspicious for  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameter. For  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters, estimation is slightly better, whilst for error variances estimation is clearly better.

The estimation for all parameters, except error variances, are poor (95 % coverage lower than .90) with all sample sizes when SMD is 1. Estimation is poor also when SMD is 2 and the sample size lower than 200, and suspicious (95 % coverage lower than .92) when the sample size is 500 for some of the parameters. Estimation is poor even when the sample size is 1000 in model A.5, B.5 and C.8. When SMD is 3, estimation is poor when the sample size is 50 or 100 in all models A.8 – C.8 and suspicious when sample size is 200 in model A.5, B.5 and C.8. When SMD is 4, estimation is poor when the sample size is 50 in all models A.8 – C.8 and suspicious when the sample size is 100 for at least one parameter in all models A.8 – C.8. When SMD is 5, estimation is poor or at least suspicious when the sample size is 50 in all models A.8 – C.8 and suspicious when the sample size is 100 for at least one parameter in all models A.8, B.8 or C.8.

As can be seen from Figures 6.44 and 6.45, the required sample sizes are in most cases largest for  $\psi_{00}$  parameter. When SMD is 3, the required sample size to achieve good estimation is between 200 – 300 and in model B.8 even 450. When SMD is 4, the required sample size to achieve good estimation is between 130 – 160 and, in model A.5, exceptionally 220. When SMD is 4, the required sample size to achieve good estimation is between 130 – 140.

sample size n

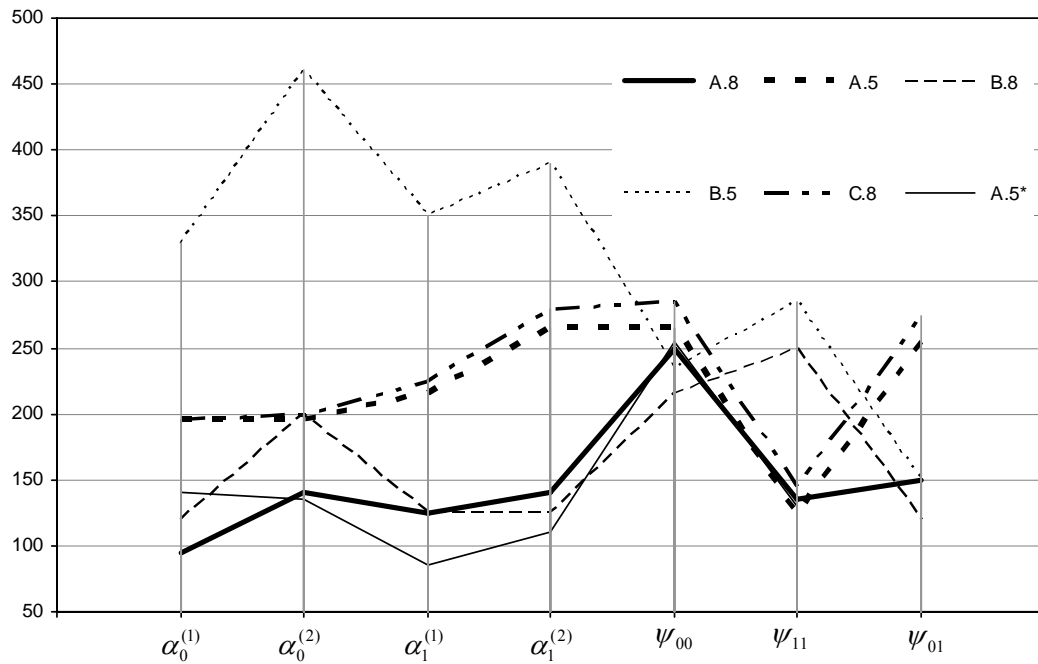


Figure 6.44. The required sample size to achieve the value .93 of the 95 % coverage, SMD=3.

sample size n

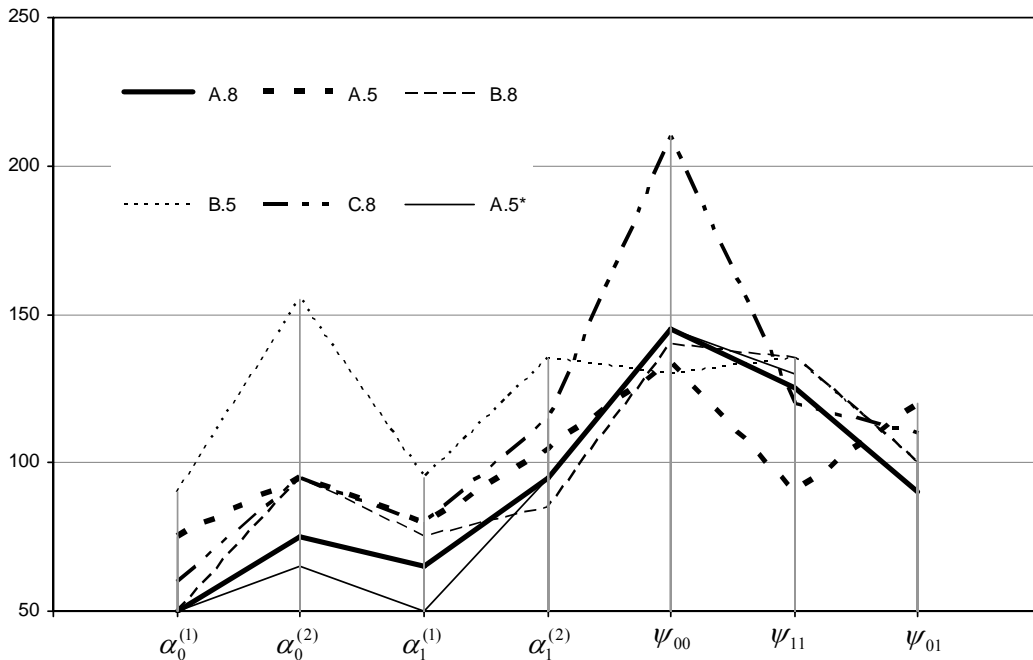


Figure 6.45. The required sample size to achieve the value .93 of the 95 % coverage, SMD=4.



### 6.3.5. Results of estimated class proportion

The next Table 6.27 shows the average of estimated class proportions ( $p^{(1)}$ ) for latent class one in the models A.8 – C.8 and standard deviation of  $p^{(1)}$  resulting of 10000 replications. In the simulation study number of observations in class one is binomially distributed  $B(n, 2/3)$ . The expected proportion  $p^{(1)}$  is .667 and when sample size is 50, 100, 200, 500 or 1000, expected standard deviations of  $p^{(1)}$  are .067, .047, .033, .021 or .015, respectively.

As in the Table 6.27 can see average proportion is very similar for each of the model A.8 – C.8. When SMD is 1, proportion is strongly downward biased increasing from .528 - .545 to .593 - .616 when sample size increase from 50 to 1000. When SMD is 2, proportion is clearly downward biased when  $n \leq 200$  and increase from .603 - .623 to .645 - .659 when sample size increase from 50 to 200 and increase to .667 - .668 when sample size increase to 1000. When SMD is 3 and sample size is 50, proportion is slightly downward biased varying between .648 and .658. Proportion is greater than .66 when SMD is 3 and sample size is greater than 100 or when SMD is 4 or 5 with all sample sizes.

Standard deviation of estimated proportion  $p^{(1)}$  is large when SMD is 1 varying between .217 and .271 in models A.8 – C.8. When SMD is 2, standard deviation of  $p^{(1)}$  is large when sample size is small but decrease strongly from .176 - .210 to .049 - .077 when sample size increase from 50 to 1000. When SMD is 3, standard deviation of  $p^{(1)}$  is between .105 - .141 in models A.8 – C.8 and is lower than .10 when  $n \geq 100$ . When SMD is 4 or 5, standard deviations of  $p^{(1)}$  are slightly greater than expected.

For summary, the bias of estimated class proportion  $p^{(1)}$  is clear when SMD is 1 or when SMD is 2 and  $n \leq 200$ . The bias is in other cases nonsignificant when the goal is to find accurate estimate of  $p^{(1)}$ . This because the standard deviation of  $p^{(1)}$  is large especially when sample size is 50 or 100. Even when standard deviation with these sample sizes are almost equal to expected value when SMD is 4 or 5, the expected standard deviation is .067 or .047, respectively. These large standard deviation should taken into account, if one is willing to estimate the near exact value of  $p^{(1)}$ .

Table 6.27. Estimated class proportion  $p^{(1)}$  for the first class and standard deviations of  $p^{(1)}$  in models A.8., A.5, A.5\*, B.8, B.5 and C.8.

Model			A8		A.5		A.5*		B.8		B.5		C.8	
n	SMD	exp. sd	p	sd	p	sd	p	sd	p	sd	p	sd	p	sd
50	1	.067	.545	.225	.537	.228	.528	.252	.545	.225	.537	.228	.536	.228
100	1	.047	.555	.229	.546	.231	.537	.263	.555	.229	.546	.231	.543	.233
200	1	.033	.570	.220	.556	.234	.548	.271	.570	.220	.556	.234	.555	.235
500	1	.021	.591	.225	.573	.236	.573	.270	.591	.225	.573	.236	.567	.238
1000	1	.015	.616	.217	.593	.232	.594	.264	.616	.217	.593	.232	.591	.233
50	2	.067	.621	.176	.604	.193	.604	.210	.623	.178	.609	.192	.603	.195
100	2	.047	.642	.151	.627	.175	.632	.190	.645	.151	.629	.175	.624	.178
200	2	.033	.656	.120	.645	.153	.653	.157	.659	.121	.648	.153	.646	.155
500	2	.021	.665	.075	.664	.111	.669	.099	.666	.076	.665	.109	.662	.113
1000	2	.015	.667	.050	.668	.077	.668	.062	.667	.049	.669	.076	.668	.079
50	3	.067	.657	.105	.651	.128	.656	.123	.658	.110	.648	.141	.649	.131
100	3	.047	.664	.072	.661	.095	.664	.084	.645	.150	.660	.108	.660	.096
200	3	.033	.665	.048	.665	.063	.665	.055	.666	.050	.666	.073	.664	.065
500	3	.021	.666	.029	.666	.037	.666	.033	.666	.031	.667	.042	.666	.037
1000	3	.015	.666	.021	.666	.026	.666	.023	.667	.022	.667	.029	.666	.026
50	4	.067	.665	.075	.663	.084	.665	.080	.665	.078	.663	.084	.662	.086
100	4	.047	.666	.051	.665	.058	.666	.054	.666	.053	.665	.058	.665	.058
200	4	.033	.666	.037	.666	.040	.666	.038	.667	.037	.666	.040	.666	.040
500	4	.021	.666	.020	.666	.025	.666	.024	.667	.024	.666	.025	.666	.025
1000	4	.015	.667	.016	.667	.018	.666	.017	.667	.017	.667	.018	.667	.018
50	5	.067	.666	.069	.665	.071	.666	.070	.666	.070	.665	.077	.666	.071
100	5	.047	.667	.048	.666	.050	.667	.049	.667	.048	.666	.053	.666	.049
200	5	.033	.667	.034	.667	.035	.667	.035	.667	.035	.667	.037	.667	.035
500	5	.021	.666	.022	.666	.023	.666	.022	.667	.022	.666	.023	.666	.022
1000	5	.015	.667	.015	.667	.017	.667	.015	.667	.015	.667	.017	.667	.016

## 7. Discussion

Despite of twenty years development of the theory behind LGM and LGMM, in terms of empirical research, these are very new methods. As these methods increasingly become more common and important to model development, the functionality of the models needs more investigation (Bauer & Curran, 2003a, 2003b; Muthén, 2004). The theory of LGM and LGMM is based on asymptotic results and, therefore, when the sample size is large, e.g. over 1000, researchers can trust the results obtained with these methods. However, in many empirical studies the sample size is small and limited to only 100-500 cases. Simulated data provides a possibility to examine the functionality of these methods with small sample sizes.

The aim of this dissertation was to examine the functionality of the LGM model, particularly with small sample sizes. The investigated model was chosen to be a linear LGMM with four repeated measurements, which is a typical case in longitudinal research. LGMM parameters were estimated using maximum likelihood estimation with robust standard errors (MLR). The functionality of LGMM was first examined with a pilot study. In this, three different constraint situations were tested with different sample sizes and different SMDs (square root of Mahalanobis distance) using linear LGM model A.8, in which the differences between two latent classes appear in the mean values of intercepts. Second, a main simulation study was carried out to examine the effect of reliability, the effect of additional measurements, and the effect of model construct, on the model estimation, with different sample sizes and different SMDs. The functionality of LGMM was approached from three different viewpoints: 1) problems in estimation of model parameters, 2) the ability of information criteria and statistical tests to decide the number of latent classes, and 3) good parameter estimation, which was evaluated using four different criteria.

The problems in estimation of model parameters was expressed as the number of failed estimations and as a number of negative variance estimates. The ability of information criteria and statistical tests to decide the number of latent classes was evaluated using three information criteria, namely AIC, BIC and aBIC, and two statistical tests, VLMR and LMR. Some results were presented concerning also the BLRT and OLRT tests. Successful parameter

estimation was evaluated using four criteria: MSE, proportion of bias in MSE, bias of standard error, and 95 % coverage.

A simulation study was carried out using the Monte Carlo method. Each of the data based on the predefined model A.8, A.5, A.5\*, B.8, B.5 or C.8 (seen in section 5.2) was replicated 10000 times to gather reliable information of LGMM functionality. Sample sizes used in simulation study were  $n = 50, 100, 200, 500$  and 1000 and differences in mean values of latent components between classes were  $SMD = 1, 2, 3, 4$  or 5, with the exception that in the pilot study  $SMD = 0.5$  was also used.

## **7.1. Results of previous studies and their limitations**

Only a few previous simulation studies concerning the latent growth mixture model have been carried out. In a simulation study, Nylund, Asparouhov, and Muthén (in press) compared the ability of different statistical indicators to resolve the number of latent classes. In this study, LGM was also one of the tested models. The results of this study were based on 100 replications. The difference between the latent classes was defined by the certain values of parameters. The examined sample sizes were  $n = 200, 500$  or 1000. The main result was that best information criteria from AIC, BIC, aBIC and CAIC was BIC. When taking account of the results of higher Type I error rate for the LMR test, BLRT was the best behaved test. Ordinary LRT, as expected, produced too high Type I error rates, which warns against using this test when deciding the number of latent classes. To conclude, the BIC index seemed to behave well and was slightly better than the BLRT for some of the models, whereas the BLRT showed the best behavior on average. For the LGM model, all the used tests did not produce satisfying results. Consequently, these results need to be more carefully examined in further simulation studies. The limitation of this study was the low number of replications, which was due to the long computing time when calculating the BLRT test value.

Only a few previous simulation studies concerning confirmatory factor mixture model have been carried out. Lubke and Muthén (2007) examined the functionality of a confirmatory factor mixture model consisting of two latent classes. This simulation study consisted of 120 replications in which the sample size was 300. The main result revealed the importance of the effect of covariate on the the 95 % coverage. However, the limitation of this simulation study is that it was limited to only one sample size and, moreover, gave information from only 95% coverage with MDs (Mahalanobis distances) 0.5, 1, and 2. The number of replications was also small.

Despite the limitations of previous simulation studies, their results provide important information about the functionality of LGMM. The mentioned limitations reveal, however, that LGMM as a method needs more investigation, which in fact makes the present simulation study well-founded and justified.

## **7.2. Conclusions based on the pilot study**

The pilot study in the present dissertation using model A.8 showed that the estimation of LGMM fails only in a few cases, and problems in estimation appear mainly in the form of negative variance estimates. Negative variance estimates seem to appear frequently in the context of more complexes (e.g., larger number of estimated parameters) LGM models, in situations I and II, and are evident particularly with small sample sizes. In practice, problems in estimation can be signs of overparameterized model. This overparameterization appears in models which include too many latent classes, or in too complex models.

Because the present simulation study concentrates on the behavior of LGMM with small sample sizes, in broader simulations (using models A.8, A.5, A.5\*, B.8, B.5 and C.8) carried out after the pilot study parameters were constrained to be equal between latent classes, as in situation III. In situation III, the number of negative variance estimates are rare enough to make it possible to examine other criteria of estimation.

## **7.3. Conclusions based on the main simulation study**

### **7.3.1. Deciding the number of latent classes**

The results of simulations suggest that when concluding the number of latent classes, AIC with all sample sizes, and aBIC with small sample sizes, are not useful because they produce too high wrong class proportions. The BIC index, which is recommended as the most reliable index in Nylund et al. (in press) simulation study, appeared to be more useful index than AIC in the present study. The results of the present study further suggest, however, that when a sample size is large, say over 500, a decision should be based on aBIC, instead of BIC. With large sample sizes, aBIC was found to be more effective than BIC in finding the right two-class solution instead of one-class

solution. In contrast to the results of Nylund et al. (in press), the present results showed that BIC is also more useful than the VLMR or LMR tests. The VLMR and LMR tests were found to be useful only with small sample sizes. When a sample size is greater than 100, they seem to produce too high wrong class proportions. When a sample size is large, 200-1000, proportions concluding to too many number of latent classes are comparable to an ordinary likelihood ratio test. Therefore, recommendation based on the results of the present study is to use BIC with small sample sizes and aBIC with large sample sizes. The borderline between small and large sample size could be about 500.

In the study of Nylund et al. (in press) BLRT showed low predictive power to find the right number of latent classes in LGMM. This result is probably due to a small SMD. Taking account of this, the few results of the present study using BLRT were found to be very promising. The results suggest that the BLRT test is useful when deciding the number of latent classes, because it produced approximately the expected proportion with a nominal .05 level when concluding to too many number of latent classes. However, because of heavy computation, only a few simulations were carried out using BLRT and, thus, more investigation is needed to further support the functionality of this test.

### **7.3.2. Failed estimation and number of negative variance estimates**

The results of main simulation study showed that estimation fails only in a few cases. However, when  $SMD \leq 3$ , the number of negative variance estimates was greater than 5 % of replications when the sample size was 50 in all tested models. When the reliability of observed variables decreases, the number of negative variance estimates shows to increase. Additional measurement points, in turn, decrease the number of negative variance estimates. If the difference in the mean value of latent components appear in the slope instead of intercept, the number of negative variance estimates decreases when reliability of observed variables are high, and increases when reliability of observed variables are low.

### **7.3.3. Results of evaluation of parameter estimation**

#### **7.3.3.1. Results of MSE**

The results concerning MSE showed that MSE decreases for all parameters in all tested models when SMD is large enough. The effect of SMD on the MSE is very strong, but the effect weakens and changes to be insignificant when SMD increases to 4. In this case, the MSE decreases by half when the sample size increases by two times. When the reliability of observed variables decreases, MSE increases strongly. Additional measurement points seem to compensate this increase in MSE. The model with correlated latent component has greater MSE compared to the model with equal SMD in which the correlation of latent components is zero. If the difference in the mean value of latent components appears in the slope instead of the intercept, MSE increases for some parameters and decreases for other parameters. These changes are more obvious when the reliability of the observed variables is low.

#### **7.3.3.2. Results of proportion of bias in MSE**

The proportion of bias in MSE was very small and appeared only for  $\alpha_0^{(1)}$ ,  $\alpha_0^{(2)}$ ,  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters. When SMD is 1, the estimates of these parameters seem to be biased. When SMD is 2, the bias for these parameters is very small, and decreases strongly when the sample size increases. These results were similar in all models A.8, A.5, A.5\*, B.8, B.5 and C.8.

#### **7.3.3.3. Results of relative bias of asymptotic standard error**

The results showed that, when SMD is 1, standard error estimates for the mean, variance and covariance parameters of latent components are badly downward biased with all sample sizes. When SMD is 2, standard errors of these parameters decrease rapidly when sample size increases. When SMD is 4 or 5, bias of standard error is lower than 5 percent with the smallest sample size ( $n = 50$ ). When the reliability of observed variables decreases, the bias of standard errors increases, requiring, for example, two times larger sample size to achieve lower than 5 percent biased standard error. Additional measurement points seem to have only a minor effect in decreasing bias. When correlation of latent components increases, bias of standard errors clearly increases. If the difference in the mean value of latent components appears in slope instead of intercept, the bias of standard errors increases for

most of the parameters, especially when the reliability of observed variables is low.

#### 7.3.3.4. Results of 95 % coverage

Because the standard error was in most cases downward biased, the 95 % coverage is lower than expected .95 value. For unbiased parameter estimation, an expected value of coverage is .937, .922 or .904 with 5, 10 or 15 % downward biased standard error, respectively. Two cut off values, .90 and .92, were chosen for the 95 % coverage to build up the results. The results suggest that if the 95 % coverage is lower than .92, estimation is suspicious and if this value is lower than .90, estimation seem to be poor. In addition to this, linear approximation is produced for sample sizes whose 95 % coverage is greater than .93. This cut off value is assumed to be a sign of good estimation.

When estimation, in terms of the 95 % coverage, is poor (95 % coverage lower than .90) or suspicious (95 % coverage lower than .92), this is seen most clearly for  $\psi_{00}$ ,  $\psi_{11}$  and  $\psi_{01}$  parameters. For  $\alpha_1^{(1)}$  and  $\alpha_1^{(2)}$  parameters, estimation seems to be almost as poor or suspicious. For  $\alpha_0^{(1)}$  and  $\alpha_0^{(2)}$  parameters, estimation is slightly better, whilst for error variances estimation is clearly better.

When SMD is 1, the estimation for all parameters, except for error variances, is poor (95 % coverage lower than .90) with all sample sizes. Estimation is poor also when SMD is 2 and the sample size lower than 200, and suspicious (95 % coverage lower than .92) for some of the parameters when the sample size is 500. In models A.5, B.5 and C.8 estimation is poor even when the sample size is 1000. When SMD is 3, estimation is poor when the sample size is 50 or 100 in all models A.8 – C.8 and suspicious when sample size is 200 in model A.5, B.5 and C.8. When SMD is 4, estimation is poor in all models A.8 – C.8 when the sample size is 50 and suspicious when the sample size is 100 for at least one parameter. When SMD is 5, estimation is poor, or at least suspicious, in all models A.8 – C.8 when the sample size is 50 and suspicious when the sample size is 100 for at least one parameter in models A.8, B.8 or C.8.

As can be seen from Figures 6.44 and 6.45, the required sample sizes to achieve a good estimation are in most cases largest for  $\psi_{00}$  parameter. When SMD is 3, the required sample size to achieve a good estimation is between 200 – 300 and, in model B.8, even 450. When SMD is 4, the required sample size to achieve good estimation is between 130 – 160 and, in model A.5,



exceptionally 220. When SMD is 4, the required sample size to achieve good estimation is between 130 – 140.

#### **7.4. Results of the simulation study that should be accounted for in empirical research**

Next, the results of this dissertation are evaluated from a practical point of view. The first issue concerns the conditions in terms of SMD and sample size within which the right number of latent classes can be decided at .70 or .80 probability. The number of latent classes is decided using BIC and aBIC as recommended previously. Then, the validity of estimation within these conditions is considered in terms of the problems in the estimation and evaluation of estimation.

The results of the simulations suggest that, when SMD is 1, it is not possible to identify the right two-class solution instead of the wrong one-class solution in any of the models A.8 – C.8. This result is thus independent of the characteristics of the model and tested sample size. When SMD is 1, MSE for mean parameters of latent components slowly decreases, the proportion of bias in MSE is relatively large, a downward bias of standard error is large, and 95% coverage is low. The MSE is large compared with the MSE in the case when SMD is 5. These results suggest that the goal to find latent classes with small effect size (SMD = 1) is unreasonable. When SMD is 1 and sample size is small, that is, 50 or 100, the number of negative variance estimates are frequent, especially when reliability of observed variables is small. The negative variance estimates are supposed to appear and are not due to the misspecification of the model.

When SMD is 2, it is possible to identify the true two-latent classes only when reliability of observed variables is high and the sample size is relatively large. When using aBIC, the probability to identify the right two-class solution instead of the wrong one-class solution is greater than .70 only in models A.8 and B.8, and the sample size needed for these models to achieve .70 - .80 probability is 600-800. If reliability is smaller (e.g., model A.5, A.5\* and B.5) or latent components correlate with each other (e.g., model C.8), even a relatively high sample size, that is,  $n=1000$ , is not enough. As in the case when SMD is 1, negative variance estimates appear frequently when SMD is 2 and the sample size is small. MSE for all parameters clearly decreases; approximately to one third when the sample size increases to 1.5-2.4 times larger. However, the MSE is large compared with the MSE in the case SMD is 5; 4-6 times larger when the sample size is large ( $n = 1000$ ). When the sample size is 600-800 and reliability of observed variables is high (models A.8 and B.8), the proportion of bias is very small and also

insignificant, the bias of standard error is lower than 5% on average, and 95% coverage is greater than .90 for all parameters. In other models the bias of standard error for some of the parameters is suspiciously high, especially when reliability is low (models A.5 and B.5), even when sample size is large  $n=1000$ . In these cases, 95% coverage is also suspiciously low, lower than .90.

When SMD is 3, it is possible to identify the right two-latent classes with small sample sizes, for which purpose BIC is most appropriate. When using BIC, the probability to identify the right two-class solution instead of the wrong one-class solution is greater than .70 in models A.8, A.5\* or B.8 when the sample size is greater than 170, 290 or 190, respectively; in models A.5 or C.8 when the sample size is greater than 365 or 375, respectively; and in model B.5 when the sample size is greater than 450. When the sample size is 500, probability for other models than model B.5 are greater than .90 if using BIC and also in model B.5 if using aBIC. MSE for all parameters clearly decreases when the sample size increases; approximately to half when the sample size increases by two times. However, the MSE is large compared with the MSE in the case SMD is 5; 1.0 – 2.5 times larger. In the cases where probability to conclude a right two-class solution is greater than .70, the proportion of bias is very small and insignificant, a bias of standard error is lower than 5% for all parameters, and 95% coverage is greater than .90 for all parameters.

When SMD is 4, the probability in identifying the right two-latent-class solution instead of the wrong one-class solution is greater than .70 with the smallest sample size ( $n=50$ ) using BIC in models A.8 and B.8. For model A.5\*, A.5, C.8 or B.5, the needed sample size to achieve .70 probability is 70, 75, 80 or 120, respectively. When the sample size is 120, probability is greater than .70 in all models, and greater than .80 in all other models, except in model B.8, and greater than .90 in A.8, B.8 and A.5\*. These results describe the strong increase in probability when the sample size increases. As in the case where SMD is 3, MSE for all parameters also clearly decreases when SMD is 4; approximately to half when the sample size increases by two times. In the cases where probability to conclude to the right two-class solution is greater than .70, the proportion of bias is very small and insignificant and the bias of standard error is greater than 5% for some parameters, but always lower than 10%. The 95% coverage is greater than .90 for all mean and error variance parameters. For variance and covariance parameters, 95% coverage is suspiciously small, slightly lower than .90, in all other models expect in model B.5. After taking account of suspicious results in bias of standard error and 95% coverage, to achieve reliable results in estimation, the sample size should be greater than presented in the context of deciding the number of latent classes at the level of .70 probability.

When SMD is 5, the proportion using BIC is greater than .90 in all models when  $n \geq 50$ . MSE for all parameters clearly decreases; approximately to half when sample size increases by two times. In the cases where  $n \geq 50$ , the proportion of bias is very small and insignificant and the bias of standard error is greater than 5% for some parameters, but always lower than 8%. A suspicious downward bias in standard errors appears in all models, most often in variances and covariance of latent components. This phenomenon is also seen as suspiciously small 95% coverage (coverage slightly lower than .90). After taking account of suspicious results in bias of standard error and 95% coverage, to achieve reliable results in estimation, the sample size should be greater than 50.

## **7.5. Implications for further studies**

The results of the present simulation study provide some challenges for further research. The results of the present dissertation are based on the linear latent growth curve mixture model with two latent classes, in which variance and covariance parameters are constrained to be equal across classes.

It could be worthwhile to analyze problems further in estimation with each simulated data. The covariation between latent components compared to the variances of latent components can be too high. Even when the estimate of covariance is nearly to exact, too low variance estimates can produce not admissible estimates of the covariance matrix of the latent components, which requires some modification to the model. Further modifications are also expected to occur because of nonsignificant parameter estimates. The important questions are; what are the expected modifications when the problems in estimation occur and what are the consequences of modifications?

Using unconstrained models probably lessens dramatically the validity of estimation and require further investigation. The results showed that the required difference between classes in terms of SMD should be very large in order to find the right number of latent classes. Previous studies have shown the meaning of covariate in finding latent classes (Lubke & Muthén, 2007). Therefore, in further simulation studies, the effect of covariate should be taken into account. Second, the results revealed that finding latent classes is dependent on the SMD of observed variables, rather than the SMD of latent variables.

Empirical distributions of the difference of likelihood ratio in model comparison suggest that the penalty seems to be very stable, slightly dependent upon the sample size. In the present study this was investigated only in the context of model A.8. It could be fruitful to investigate this phenomenon in other models as well.

If distribution of observed variables is skewed, overextraction of latent classes is obvious (Bauer & Curran, 2003a, 2003b; Cudeck & Henly, 2003; Muthén, 2003; Rindskopf, 2003), and, therefore, one should be aware of the effect of skewed distribution when using LGMM in practice. This phenomenon set also future challenges to examine LGMM with skewed distribution.

The mixture modeling can be extended to various models, for example, factor mixture model and latent variable hybrids (Muthén, 2006). When using mixture modeling in empirical data, it is important to examine the behavior of various different models with small sample sizes, as was done in this work for LGMM. This requires a lot of simulation studies.

## References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-332.
- Anstey, K. J., Hofer, S. M., & Luszcz, M. A. (2003). A latent growth curve analysis of latent-life sensory and cognitive function over 8 years: Evidence for specific and common factors underlying change. *Psychology and Aging*, 18(4), 714-726.
- Arbuckle, J. L. (2006). *Amos 7.0 User's Guide*. Chicago: SPSS.
- Aunola, K., Leskinen, E., Onatsu-Arvilommi, T., & Nurmi, J.-E. (2002). Three methods for studying developmental change: A case of reading skills and self-concept. *British Journal of Educational Psychology*, 72, 343-364.
- Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to Grade 2. *Journal of Educational Psychology*, 96, 699-713.
- Bauer, D. J., & Curran, P. J. (2003a). Overextraction of Latent Trajectory Classes: Much Ado About Nothing? Reply to Rindskopf (2003), Muthen (2003), and Cudeck and Henly (2003). *Psychological Methods*, 2( 3), 384-393.
- Bauer, D. J., & Curran, P. J. (2003b). Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes. *Psychological Methods*, 8(3), 338-363.
- Bauer, D. J., & Curran, P. J. (in press). The Behaviour of Growth Mixture Models Under Nonnormality: A Monte Carlo Analysis. *Psychological Methods*.
- Beauducel A., Herzberg P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least square estimation in CFA. *Structural Equation Modelling*, 13(2), 186-203.
- Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software.
- Bentler, P. M., & Yuan, K-H. (1999). Structural equation modelling with small samples: Test statistics. *Multivariate Behavioral Research*, 34(2), 181-197.
- Bergman, L. R., Magnusson, D., & El-Khoury, B. M. (2003). Studying individual development: A person-oriented approach. In D. Magnusson (Series Ed.), *Paths through life* (Vol. 4). Mahwah, NJ: Lawrence Erlbaum Associates.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*(1), 30-52.

Bollen K. A., & Curran, P.J. (2006). *Latent Curve Models. A structural Equation Perspective*. Hoboken, New Jersey: John Wiley & Sons.

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modelling revisited. In R. Cudeck, S. du Toit, D. Sörbom (Ed.), *Structural Equation Modeling: Present and Future (pp. 139-164)*. Scientific Software International.

Bryk, A. S., & Raudenbush S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*(1), 147-158.

Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models*. Sage publications, Inc. Newbury Park, California.

Chan, D., Ramey, S., Ramey, C., & Schmitt, N. (2000). Modeling intraindividual changes in children's social skills at home and at school: A multivariate latent growth approach to understanding between-settings differences in children's social skill development. *Multivariate Behavioral Research, 35*(3), 365-396.

Chen, F., Bollen, K., Paxton, P., Curran, P. J., & Kirby, J. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research, 29*, 468-508. Accompanying Technical Appendix for Chen et al. (2001).

Chou, C.-P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curve: The multilevel model and the latent curve analysis. *Structural equation modelling, 5*(3), 247-266.

Colder, C. R., Mehta, P., Balanda, K., Campbell, R. T., Mayhew, K. P., Stanton, W. R., Pentz, M-A., & Flay, B. R. (2001). Identifying trajectories of adolescent smoking: An application of latent growth mixture modeling. *Health Psychology, 20*(2), 127-135,

Cudeck, R., Henly, S. J. (2003). A Realistic Perspective on Pattern Representation in Growth Data: Comment on Bauer and Curran (2003). *Psychological Methods, 8*(3), 378-383.

Curran, P. J., & Hussong, A. M. (2003). The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology, 112*(4), 526-544.

Duncan, T. E., Duncan, S. C., Alpert, A., Hops, H., Stoolmiller M., & Muthén, B. (1997). Latent variable modeling of longitudinal and multilevel substance use data. *Multivariate Behavioral Research*, 32(3), 275-318.

Duncan, T. E., Duncan S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variables growth curve modeling*. Mahwah, New Jersey, London.

Duncan, T. E., & Tildesley, E. (1995). The Consistency of family and peer influences on the development of substance use in adolescence. *Addiction*, 90(12), 1647-1660.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modelling fit indexes. *Structural Equation Modeling*, 6(1), 56-83.

Fuzhong, L., Barrera, M., Hops H., Fisher. K. J. (2002). The longitudinal influence of peers on the development of alcohol use in late adolescence: A growth mixture analysis. *Journal of Behavioral Medicine*, 25(3)293-315.

Goldstein, Harvey (1987). *Multilevel models in educational and social research*. Griffin. London.

Goldstein, Harvey (1996). *Multilevel statistical models*. Arnold. London.

Hancock, G. R., Kuo, W.-L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, 8(3), 470-489.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, 11(1), 36-53.

Hoogland J.J., Boomsma A. (1998). Robustness studies in covariance structure modelling. An overview and a meta-analysis. *Sociological Methods & Research*, 26(3), 329-367.

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In Hoyle, R. H. (Ed.), *Structural equation modelling: Concepts, issues, and applications* (pp. 75-99). London, New Delhi: SAGE publications.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.

Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling*, 8(2), 205-223.

Jones, B., Nagin, D., & Roeder, K., (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods and Research*, 29, 374-393.

Jöreskog, K. G., Sörbom, D., Du Toit, S., & Du Toit, M (1999). *LISREL 8: New Statistical Features*. Chicago: Scientific Software International.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86.

Li, F., Barrera, Jr., M., Hops, H., & Fisher, K. J. (2002). The longitudinal influence of peers on the development of alcohol use in late adolescence: A growth mixture analysis. *Journal of Behavioral Medicine*, 25(3), 293-315.

Lo, Y., Mendell, R. N., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778.

Longford, Nicholas T. (1993). Random coefficient models. Clarendon Press. Oxford.

Lubke, G., Muthén B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods* Vol.10, No. 1, 21-39,

Lubke, G., & Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects and class-specific parameters. *Structural Equation modelling: A Multidisciplinary Journal*. 14(1), 26-47.

Lyytinen, H., Tolvanen, A., Torppa, M., Poikkeus, A.-M., & Erskine J. (2006). Trajectories of reading development: A follow-up from birth to school age of children with and without risk for dyslexia. *Merrill-Palmer Quarterly*, 52(3), 514-546.

MacCallum, R. C., Kim, C., Malarkey, W.B., & Kiecolt-Glaser, J. K. (1997). Studying Multivariate Change Using Multilevel Models and Latent Curve Models. *Multivariate Behavioral Research*, 32(3), 215-253.

MacCulloch, C.E., Searle S. R. (2001). Generalized, linear, and mixed models. New York: Wiley & Sons.



Mason, W. A. (2001). Self-esteem and delinquency revisited (again): A test of Kaplan's self-derogation theory of delinquency using latent growth curve modelling. *Journal of Youth and Adolescence*, 30(1), 83-102.

McArdle, J. J., & Aber, M. S. (1990). Patterns of change within latent variable structural equation models. In A. Von Eye (Ed.), *Statistical methods in longitudinal research: Vol. I, Principles and structuring change* (pp. 151-224). New York: Academic Press.

McArdle, J. J., & Epstein D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110-133.

McArdle J. J., Nesselroade J. R. (2002). Growth curve analysis in contemporary psychological research. In J. Schinka & W. Velicer (Ed.), *Comprehensive handbook of psychology, Volume Two: Research Methods in Psychology* (pp. 447-480). New York: Wiley.

McLahlan, G. (1999). Mahalanobis Distance. *Resonance*, 4, 20-26.

McLahlan, G.J., & Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: John Wiley and Sons.

McLahlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley & Sons.

Meredith, W., & Tisak J. (1990). Latent curve analysis. *Psychometrika*, 55, 107-122.

Muthén, B. O. (2001). Latent Variable Mixture Modelling. In G. A. Marcoulides & R. Schumacker (Eds.), *New developments and techniques in structural equation modelling* (pp. 1-33). Mahwah, Lawrence Erlbaum Associates.

Muthén, B. (2003). Statistical and Substantive Checking in Growth Mixture Modeling: Comment on Bauer and Curran (2003). *Psychological Methods*, 8(3), 369-377.

Muthén, B. (2004). Latent variable analysis: Growth mixture modelling and related techniques for longitudinal data. In D. Kaplan (ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthén, B. (2006). Latent variable hybrids: *Overview of old and new models. Forthcoming in Hancock, G. R., & Samuelsen, K. M. (Eds.). (2007). Advances in latent variable mixture models*. Charlotte, NC: Information Age Publishing.

Muthén, B., Khoo, S-T., Francis, D. J., & Boscardin, C. K. (2003). Analysis of reading skills development from kindergarten through first grade: An application of growth mixture modelling to sequential processes. In S. R. Reise & N. Duan (Eds.), *Multilevel modelling: Methodological advances, issues and applications* (pp. 71-89). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, L. K., & Muthen, B. O. (1998-2006). *Mplus User's Guide*. LA: Muthén & Muthén.

Muthén, L. K. & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 4, 599-620.

Muthén, B., & Shedden, K. (1999). Finite mixture modelling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463-469.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, 4, 139-157.

Neale MC, Boker SM, Xie G, Maes HH (1999). *Mx: Statistical modelling*. Box 126 MCV, Richmond, VA 23298: Department of Psychiatry. 5<sup>th</sup> Edition.

Nylund, K. L., Asparouhov, T., & Muthen, B. (in press). Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study.

Olsson, U. H., Troye, S. V., & Howell, R. D. (1999). Theoretical fit and empirical fit: The performance of maximum likelihood versus generalized least squares estimation in structural equation models. *Multivariate Behavioral Research*, 34(1), 31-58.

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology*, 97(3), 299-319.

Paxton, P. M., Curran, P., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo Experiments: Design and Implementation. *Structural Equation Modeling* 8, 287-312.

Powell, D. A., Schafer, W. D. (2001). The robustness of the likelihood ratio chi-square test for structural equation models: A meta analysis. *Journal of Educational and Behavioral statistics*, 26(1), 105-132.

Reynolds, C. A., Finkel, D., McArdle, J. J., Gatz, M., Berg, S., & Pedersen, N. L. (2005). Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. *Developmental Psychology*, *41*(1), 3-16.

Rindskopf, D. (2003). Mixture or Homogeneous? Comment on Bauer and Curran (2003). *Psychological Methods*, *8*(3), 364-368.

Rovine, M. J., & Molenaar, P. C. M. (1998). A nonstandard method for estimating a linear growth model. *International Journal of Behavior and Development*, *22*(3), 453-473.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333-343.

Tolvanen, A. (2000). *Latenttien kasvukäyrä- ja simplex-mallien teoriaa ja sovelluksia pitkittäisaineistoissa kehityksen ja muutoksen analysointiin*. Licentiate thesis, Department of Statistics, University of Jyväskylä.

Torppa M., Tolvanen A., Poikkeus A.-M., Eklund K., Lerkkanen M.-K., Leskinen E, & Lyytinen H. (in press). Reading development subtypes and their early characteristics. *Annals of Dyslexia*.

van Lier, P. A. C., Muthén, B. O., van der Sar, R. M., & Crijnen, A. A. M. (2004). Preventing disruptive behaviour in elementary schoolchildren: Impact of a universal classroom-based intervention. *Journal of Consulting and Clinical Psychology*, *72*(3), 467-478.

White, H., (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1-25.

Wickrama, K. A. S., Lorenz F. O., Conger R.D., Glen H.E (1997). Marital quality and physical illness: A latent growth curve analysis. *Journal of Marriage & the Family*, *59*(1)143-155.

Windle, M., Windle. R. C. (2001). Depressive symptoms and cigarette smoking middle adolescents: Prospective associations and intrapersonal and interpersonal influences. *Journal of Consulting and Clinical Psychology*, *69*(1)215-226.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, *57*( 2), 307-333.

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

Appendix 1. Mplus script to generate and analyze data by using parameters of model A.8 with  $SMD = 3$  and  $n = 500$ .

```
TITLE:          LGM A.8 model SMD=3 n=500
MONTECARLO:    NAMES ARE x1 x2 x3 x4;
               NOBSERVATIONS = 500;
               NREPS = 10000;
               SEED =53648292;
               GENCLASSES = C(3);
               CLASSES = C(2);
               SAVE = SMD3A8n500.SAV ;
               RESULTS = SMD3A8n500.RES;
ANALYSIS:      TYPE = MIXTURE;
               ESTIMATOR = MLR;
               MITERATIONS = 1000;
MODEL MONTECARLO:
               %OVERALL%
               i BY x1-x4@1;
               s BY x1@0 x2@1 x3@2 x4@3;
               [x1-x4@0];
               [i*0 s*.2];
               i*1.0;
               s*0.2;
               i with s*0;
               x1*.25 x2*.30 x3*.45 x4*.70;
               %C#1%
               [i*0 s*.2];
               %C#2%
               [i*0 s*.2];
               %C#3%
               [i*3 s*.2];
MODEL:
               %OVERALL%
               i BY x1-x4@1;
               s BY x1@0 x2@1 x3@2 x4@3;
               [x1-x4@0];
               [i*0 s*.2];
               i*1;
               s*0.2;
               i with s*0;
               x1*.25 x2*.30 x3*.45 x4*.70;
               %C#1%
               [i*0 s*.2];
               %C#2%
               [i*3 s*.2];
OUTPUT:       TECH9 TECH11;
```

Appendix 2. Calculating difference of information criteria and statistical tests to decide number of latent classes – two class solution versus one class solution. The data is produced using Mplus script presented in Appendix 1 and cleaned up using script in text editor.

```
data list file='d:\LGMA8\SMD3A8n500.txt' Free /
repro logli lo2dif difdf mea std p1 value p2.
execute.

compute nobs=500.
compute xpvalue=1-cdf.chisq(lo2dif,difdf).
recode xpvalue
(0 thru .0000999999=0)(0.0001 thru .0009999999=1)
(0.001 thru .0099999999=2)(.01 thru .0499999999=3)
(.05 thru .0999999999=4)(.10 thru hi=5) into xp05.

value labels xp05 0 'p<.0001' 1'p<.001' 2 'p<.01'
3 'p<.05' 4 'p<.10' 5 'ns.'.

var labels xp05 'ordinary LRT'.
execute.

recode p1 p2 (0 thru .0000999999=0)(0.0001
thru .0009999999=1)
(0.01 thru .0099999999=2)(.01 thru .0499999999=3)
(0.02 (.05 thru .0999999999=4)(.10 thru hi=5) into p105
p205.

value labels p105 p205 0 'p<.0001' 1'p<.001' 2 'p<.01'
3 'p<.05' 4 'p<.10' 5 'ns.'.

var labels
p105 'Lo-Mendel-Rubin test'
p205 'Adjusted Lo-Mendel-Rubin test'.
execute.

frequencies variables=p105 p205 xp05.

compute AICERO=-1*lo2dif+2*difdf.
compute BICERO=-1*lo2dif+ln(nobs)*difdf.
compute aBICERO=-1*lo2dif+ln((nobs+2)/24)*difdf.

var label
aicero 'AIC2-AIC1'
bicero 'BIC2-BIC1'
```

```
abicero 'aBIC2-aBIC1'  
p1 'LO-.. p-value'  
p2 'LO-...adj p-value'.  
execute.
```

```
recode AICERO BICERO aBICERO (lo thru -.00000001=2)(0  
thru hi=1).  
execute.  
freq var=AICERO BICERO aBICERO.
```

Appendix 3. Calculating number of negative variance estimates in all and for each variance parameter separately. The data is produced using Mplus script presented in Appendix 1.

```
data list file='d:\LGMA8\SMDA8n500.res ' Free /
repro converge loglikel
tel te2 te3 te4 ilalpha slalpha ilpsi islpsi slpsi
i2alpha s2alpha calpha.
execute.
```

```
freq konverg.
```

```
select if (konverg eq 0).
execute.
```

```
*number of negative variances *
```

```
if (konverg = 0) negvar=0.
if (tel lt 0) negvar =1.
if (te2 lt 0) negvar =1.
if (te3 lt 0) negvar =1.
if (te4 lt 0) negvar =1.
if (ilpsi lt 0) negvar =1.
if (slpsi lt 0) negvar =1.
execute.
if (konverg ge 1) negvar =2.
```

```
value labels negvar 0 'no problems' 1 'negative
variances'
2 'failed estimation'.
```

```
FREQUENCIES VARIABLES=negvar /ORDER= ANALYSIS .
```

```
recode tel te2 te3 te4 ilpsi slpsi
(lo thru -.00000000001=0)(0 thru hi=1).
freq var=tel te2 te3 te4 ilpsi slpsi.
```



76. HEIKKOLA, ERKKI, Domain decomposition method with nonmatching grids for acoustic scattering problems. (94 pp.) 1997
77. TOIVANEN, JARI, Fictitious domain method applied to shape optimization. (108 pp.) 1997
78. SAARIMÄKI, MIKKO, Disjointness and complementedness in upper continuous lattices. (7 pp.) 1998
79. HINKKANEN, AIMO, Applications of complex dynamics and potential theory. (104 pp.) 1998
80. IWANIEC, TADEUSZ, Nonlinear differential forms. (207 pp.) 1998
81. KAUKANEN, JANNE, Condition N for Sobolev mappings. (18 pp.) 2001
82. TASKINEN, ILKKA, Cluster priors in the Bayesian modelling of fMRI data. (105 pp.) 2001
83. PAPERS ON ANALYSIS: A VOLUME DEDICATED TO OLLI MARTIO ON THE OCCASION OF HIS 60TH BIRTHDAY. Edited by J. Heinonen, T. Kilpeläinen, and P. Koskela. (315 pp.) 2001
84. ONNINEN, JANI, Mappings of finite distortion: Continuity. (24 pp.) 2002
85. OLLILA, ESA, Sign and rank covariance matrices with applications to multivariate analysis. (42 pp.) 2002
86. KAUKO, VIRPI, Visible and nonexistent trees of Mandelbrot sets. (26 pp.) 2003
87. LLORENTE, JOSÉ G., Discrete martingales and applications to analysis. (40 pp.) 2002
88. MITSIS, THEMIS, Topics in harmonic analysis. (52 pp.) 2003
89. KÄRKKÄINEN, SALME, Orientation analysis of stochastic fibre systems with an application to paper research. (53 pp.) 2003
90. HEINONEN, JUHA, Geometric embeddings of metric spaces. (44 pp.) 2003
91. RAJALA, KAI, Mappings of finite distortion: Removable singularities. (23 pp.) 2003
92. FUTURE TRENDS IN GEOMETRIC FUNCTION THEORY. RNC WORKSHOP JYVÄSKYLÄ 2003. Edited by D. Herron. (262 pp.) 2003
93. KÄENMÄKI, ANTTI, Iterated function systems: Natural measure and local structure. (14 pp.) 2003
94. TASKINEN, SARA, On nonparametric tests of independence and robust canonical correlation analysis. (44 pp.) 2003
95. KOKKI, ESA, Spatial small area analyses of disease risk around sources of environmental pollution: Modelling tools for a system using high resolution register data. (72 pp.) 2004
96. HITCZENKO, PAWEŁ, Probabilistic analysis of sorting algorithms. (71 pp.) 2004
97. NIEMINEN, TOMI, Growth of the quasihyperbolic metric and size of the boundary. (16 pp.) 2005
98. HÄHLOMAA, IMMO, Menger curvature and Lipschitz parametrizations in metric spaces. (8 pp.) 2005
99. MOLTCHANOVA, ELENA, Application of Bayesian spatial methods in health and population studies using registry data. (55 pp.) 2005
100. HEINONEN, JUHA, Lectures on Lipschitz analysis. (77 pp.) 2005
101. HUJO, MIKA, On the approximation of stochastic integrals. (19 pp.) 2005
102. LINDQVIST, PETER, Notes on the  $p$ -Laplace equation. (80 pp.) 2006
103. HUKKANEN, TONI, Renormalized solutions on quasi open sets with nonhomogeneous boundary values. (41 pp.) 2006
104. HÄHKIÖNIEMI, MARKUS, The role of representations in learning the derivative. (101 pp.) 2006
105. HEIKKINEN, TONI, Self-improving properties of generalized Orlicz–Poincaré inequalities. (15 pp.) 2006
106. TOLONEN, TAPANI, On different ways of constructing relevant invariant measures. (13 pp.) 2007
107. HORPPU, ISMO, Analysis and evaluation of cell imputation. (248 pp.) 2008
108. SIRKIÄ, SEIJA, Spatial sign and rank based scatter matrices with applications. (29 pp.) 2007
109. LEIKAS, MIKA, Projected measures on manifolds and incomplete projection families. (16 pp.) 2007
110. TAKKINEN, JUHANI, Mappings of finite distortion: Formation of cusps. (10 pp.) 2007