Bayesian Criterion-based Model Selection in Structural Equation Models

LI, Yunxian

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Statistics

The Chinese University of Hong Kong
June 2010

Thesis/Assessment Committee

Professor LEE Sik-Yum (Chair)

Professor SONG Xin-Yuan (Thesis Supervisor)

Professor CHAN Ping-Shing (Committee Member)

Professor TANG Man-Lai (External Examiner)

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree of qualification of this or any other university or other institution of learning.

Abstract of thesis entitled:

Bayesian Criterion-based Model Selection in Structural Equation Models

Submitted by LI, Yunxian

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in May 2010

Structural equation models (SEMs) are commonly used in behavioral, educational, medical, and social sciences. Lots of software, such as EQS, LISREL, MPlus, and WinBUGS, can be used for the analysis of SEMs. Also many methods have been developed to analyze SEMs. One popular method is the Bayesian approach. An important issue in the Bayesian analysis of SEMs is model selection. In the literature, Bayes factor and deviance information criterion (DIC) are commonly used statistics for Bayesian model selection. However, as commented in Chen et al. (2004), Bayes factor relies on posterior model probabilities, in which proper prior distributions are needed. And specifying prior distributions for all models under consideration is usually a challenging task, in particular when the model space is large. In addition, it is well known that Bayes factor and posterior model probability are generally sensitive to the choice of the prior distributions of the parameters. Furthermore the computational burden of Bayes factor is heavy. Alternatively, criterion-based methods are attractive in the sense that they do not require proper prior distributions in general, and the computation is quite simple. One of commonly used criterion-based methods is DIC, which however assumes the posterior mean to be a good estimator. For some models like the mixture SEMs, WinBUGS does not provide the DIC values. Moreover, if the difference in DIC values is small,

only reporting the model with the smallest DIC value may be misleading. In this thesis, motivated by the above limitations of the Bayes factor and DIC, a Bayesian model selection criterion called the L_v measure is considered. It is a combination of the posterior predictive variance and bias, and can be viewed as a Bayesian goodness-of-fit statistic. The calibration distribution of the L_v measure, defined as the prior predictive distribution of the difference between the L_v measures of the candidate model and the criterion minimizing model, is discussed to help understanding the L_v measure in detail. The computation of the L_v measure is quite simple, and the performance is satisfactory. Thus, it is an attractive model selection statistic. In this thesis, the application of the L_v measure to various kinds of SEMs will be studied, and some illustrative examples will be conducted to evaluate the performance of the L_v measure for model selection of SEMs. To compare different model selection methods, Bayes factor and DIC will also be computed. Moreover, different prior inputs and sample sizes are considered to check the impact of the prior information and sample size on the performance of the L_v measure. In this thesis, when the performances of two models are similar, the simpler one is selected.

摘要

結構方程模型(SEMs)被廣泛應用於行為學、教育學、醫學以及社會科學等領域。很多軟件都可以用來分析SEMs，比如 EQS, LISREL, MPlus, WinBUGS，等等。同時，很多统计方法可以用于分析SEMs。在貝葉斯模型選擇中，貝葉斯因子和DIC準則是其中最常用的兩種方法。模型選擇在結構方程模型分析中是一個重要的問題。在一些文獻中，貝葉斯因子被廣泛運用于結構方程模型的比較 。然而，這種方法通常依賴於模型的後驗概率，从而也依赖于合適的先驗分布的选取。在可選擇模型比較多的時候，這通常是一項很艰巨的任務。除此之外，貝葉斯因子以及後驗概率通常受參數先驗分布的影響。而且，貝葉斯因子的計算量較大。然而，基於準則的方法通常不需要选择適當的先驗分布，而且計算也比較簡單。最常用的模型選擇準則之一是DIC準則，然而DIC準則通常假設後驗均值是較优的估計。對於一些模型，例如混合模型，WinBUGS不能提供DIC值。另外，如果兩個模型的DIC值相差很小，只考虑具有最小DIC值的那个模型可能会得出错误的结果。基于以上提到的关于贝叶斯因子和DIC准则的一些缺点，我們提出了一種叫做Lv 測度的貝葉斯準則作為模型選擇的方法。Lv測度是後驗預測方差和偏差的一個組合，也可以看成一個貝葉斯擬合優度統計量。同時我們還定義了Lv測度的一個校正分布，即候選模型與最小Lv測度模型的Lv測度之差的邊際分布，這將作為Lv測度的一個補充。正如Ibrahim et al.(2001)指出，Lv測度的計算比較簡單，而且在很多模型的选择中有较好的表现。因此，這是一個很有吸引力的模型選擇統計量 。本論文中，我們將討論Lv測度在結構方程模型中的應用，同時給出一些實例對這種方法在結構方程模型選擇中的表現進行評估。為了比較不同的模型選擇方法，我們也給出貝葉斯因子和DIC準則的相关結果。此外，我們還討論了先驗信息及樣本量對Lv測度的影響。在本文中，對於表現相同的模型，我們將選擇相對簡單的模型。

Acknowledgement

I would like to thank, Prof. Song Xin-Yuan, my supervisor, for her many suggestions and constant support during the period of my PhD program. I am also thankful to Prof. Lee Sik-Yum for his guidance through the early years of my PhD research. Special thanks to Prof. Chan Ping-Shing, Prof. Lee Sik-Yum, and Prof. Tang Man-Lai for their service as committee members and giving valuable comments on my research. It is also pleasure to express my gratitude to all staff of the Department of Statistics, The Chinese University of Hong Kong for their kind assistance. And extra special thanks to my family and friends for the love, support and encouragement over the years.

Contents

Chapter 1

Introduction

1.1 Overview

Structural equation models (SEMs) are very popular in analyzing relationships among observed and latent variables. Nowadays, SEMs have been applied to many fields, including but not limited to business, marketing, education, medicine, psychology and sociology. One of the main objectives of these applications is to search a good SEM that can reveal the relationships among covariates, observed and latent variables. Hence, model selection is an important issue in analyzing SEMs. Moreover, as explained in Lee (2007, chap. 5), hypothesis testing can be treated as a model selection problem.

Recently, the Bayesian approach for analyzing SEMs has received much attention, see Schines et al. (1999), Dunson (2000), Ansari et al. (2002), Lee (2007), Lee et al. (2010), Cai et al. (2010), and references therein. So far, the most widely used Bayesian model selection statistics are Bayes factor (Kass and Raftery, 1995) and the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). It is well known that for complex statistical models, the computation of Bayes factor is difficult (DiCiccio et al., 1997). Gelman and Meng (1998) developed an efficient algorithm, namely the path sampling, to compute the normalizing constant of a probability density function. This algorithm has been applied to compute the Bayes factors of many complex

SEMs (see for example, Lee and Song, 2002, 2003a; Song and Lee, 2007, 2008, and the references therein). Like the Bayesian Information Criterion (BIC), DIC takes into account the number of unknown parameters in the model. As the software WinBUGS (Spiegelhalter et al., 2003) provides the DIC values for most SEMs, the application of DIC is convenient.

While Bayes factor and DIC have some nice features, they have limitations. It is well-known that Bayes factor requires proper prior distributions of the parameters. In fact, it will favor the competitive model M_0 if the prior of the parameters in model M_1 has a very large spread so as make it non-informative. This is known as the "Bartletts Paradox". Moreover, for competitive models M_0 and M_1, such as multilevel SEMs with very different structures, it is difficult to find a direct path to link them when applying the path sampling. Under these cases, some auxiliary models may have to be used in computing the Bayes factor (see Lee, 2007). This will increase the computational burden. For DIC, it assumes the posterior mean to be a good estimator; and for some models (for example, the mixture SEMs), WinBUGS does not give the DIC values. Moreover, if the difference in DIC values is small, only reporting the model with the smallest DIC value may be misleading. In this thesis, motivated by the above limitations of the Bayes factor and DIC, we propose an attractive Bayesian statistic for model selection for different kinds of SEMs.

The proposed Bayesian statistic, called the L_v measure, is a criterion-based method that does not require proper prior distributions of the parameters. It will be shown that the computational burden involved is light, and the statistic can be obtained conveniently via observations simulated for the Bayesian estimation. Basically, the L_v measure involves two components. The first component is related to the reliability of the prediction, and the second component measures the discrepancy between the prediction and the observed data. Hence, it can be used to examine the goodness-of-fit of the model to the observed data. We will also consider the calibration dis-

tribution of the L_v measure, which will allow us to compare two competing models in more details.

1.2 L_v Measure for Model Selection

To define the criterion, some notations will be used. Let $\boldsymbol{Y}^{obs} = (\boldsymbol{y}_1^{obs}, \cdots, \boldsymbol{y}_n^{obs})$ be a matrix of observations, $\boldsymbol{Y}^{rep} = (\boldsymbol{y}_1^{rep}, \cdots, \boldsymbol{y}_n^{rep})$ be a matrix of replications, which has the same distribution with $\boldsymbol{Y}^{obs}$, and $\boldsymbol{\theta}$ be a vector that contains all the unknown parameters in the given model.

Gelfand and Ghosh (1998) proposed a minimum posterior predictive loss approach for model selection. They obtained the criterion by minimizing posterior loss for a given model and then, for models under consideration, selecting the one which minimizes this criterion. To define the loss, let $\boldsymbol{a} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ be the action matrix which is an estimate trying to accommodate both $\boldsymbol{Y}^{obs}$, and what we predict for $\boldsymbol{Y}^{rep}$. For the ith observation in $\boldsymbol{Y}^{obs}$, let $L(\boldsymbol{y}_i^{rep}, \boldsymbol{a}_i; \boldsymbol{y}_i^{obs})$ denote the loss for guessing $\boldsymbol{a}_i$ when $\boldsymbol{y}_i^{rep}$ is obtained and $\boldsymbol{y}_i^{obs}$ is observed. The criterion is defined as follows:

$$\begin{aligned} L(\boldsymbol{Y}^{rep}, \boldsymbol{a}; \boldsymbol{Y}^{obs}) &= \sum_{i=1}^{n} E_{\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}} L(\boldsymbol{y}_i^{rep}, \boldsymbol{a}_i; \boldsymbol{y}_i^{obs}) \\ &= \sum_{i=1}^{n} [E_{\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}} L(\boldsymbol{y}_i^{rep}, \boldsymbol{a}_i) + kL(\boldsymbol{y}_i^{obs}, \boldsymbol{a}_i)], \end{aligned} \tag{1.1}$$

where $L(\cdot, \cdot)$ denotes a certain loss, and different loss will give different criterion for model choice. $E_{\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}} L(\cdot, \cdot)$ indicates the conditional expectation of $L(\cdot, \cdot)$ which is taken with respect to the conditional distribution $p(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs})$. This equation rewards closeness to $\boldsymbol{y}_i^{rep}$ but also to $\boldsymbol{y}_i^{obs}$, $\boldsymbol{a}_i$ is viewed as a compromise action. The domain $\mathcal{A}$ for $\boldsymbol{a}_i$ need not to concur with the support of $\boldsymbol{y}_i$. For instance, if $\boldsymbol{y}_i$ is a p-dimensional vector of discrete data, say p Poisson variables, $\mathcal{A}$ would be $\underbrace{R^+ \otimes \cdots \otimes R^+}_{p}$. When the mean of $\boldsymbol{y}_i$ exists, $\mathcal{A}$ will typically be the space of the mean. The scaler k in equation (1.1) indicates the relative regret for departure from $\boldsymbol{Y}^{obs}$ as

compared with departure from $\boldsymbol{Y}^{rep}$. When $k = 0$, this criterion can be viewed as finding a better guess $\boldsymbol{a}$ for $\boldsymbol{Y}^{rep}$. By using the expected squared Euclidean distance (Ibrahim and Laud, 1994) as the loss function $L(\cdot, \cdot)$, the criterion can be defined as

$$\begin{aligned} L(\boldsymbol{Y}^{rep}, \boldsymbol{a}; \boldsymbol{Y}^{obs}) &= \sum_{i=1}^{n} E_{\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}} L(\boldsymbol{y}_i^{rep}, \boldsymbol{a}_i; \boldsymbol{y}_i^{obs}) \\ &= \sum_{i=1}^{n} E_{\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}} (\boldsymbol{y}_i^{rep} - \boldsymbol{a}_i)^T (\boldsymbol{y}_i^{rep} - \boldsymbol{a}_i) \qquad (1.2) \\ &+ \sum_{i=1}^{n} k(\boldsymbol{y}_i^{obs} - \boldsymbol{a}_i)^T (\boldsymbol{y}_i^{obs} - \boldsymbol{a}_i). \end{aligned}$$

The minimizing $\boldsymbol{a}_i$ is $(1+k)^{-1}(\boldsymbol{\mu}_i + k\boldsymbol{y}_i^{obs})$, where $\boldsymbol{\mu}_i = E(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs})$. Inserting these $\boldsymbol{a}_i$ into equation (1.2), and let $v = k/(k+1)$, we get the L_v measure (Ibrahim et al., 2001), which is given by

$$\begin{aligned} L_v(\boldsymbol{Y}^{obs}) &= \sum_{i=1}^{n} [E_{\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}} (\boldsymbol{y}_i^{rep} - \boldsymbol{\mu}_i)^T (\boldsymbol{y}_i^{rep} - \boldsymbol{\mu}_i) + v(\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs})^T (\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs})] \\ &= \sum_{i=1}^{n} \left\{ \mathrm{tr}[Var(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs})] + v(\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs})^T (\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs}) \right\}. \qquad (1.3) \end{aligned}$$

From the definition, the L_v measure can be viewed as a combination of two terms. The first one is the predicted variance which can be viewed as a penalty, and the second one is the predicted bias which can be viewed as a goodness-of-fit measure. Therefore, the model with the smallest value of the L_v measure will be selected.

1.3 Outline of the Thesis

In this thesis, we focus on model selection for several different kinds of SEMs. To the best of our knowledge, Bayes factor and DIC are the most popular methods for model selection of SEMs. Due to the reasons given in the section of overview, they have some limitations. Hence, there is a need to develop an efficient and simple approach to deal with the problem of

model selection in structural equation modeling. In Chapter 2, the L_v measure is applied to nonlinear SEMs. In addition, the calibration distribution of the L_v measure is discussed. In Chapter 3, the L_v measure is further applied to nonlinear SEMs with mixed continuous and ordinal categorical responses. In Chapter 4, considering the existence of hierarchical observations in real applications, the L_v measure, together with the calibration distribution, is used for model selection of two-level SEMs. In Chapter 5, a finite mixture of SEMs with unknown number of components is considered for the analysis of heterogeneous data. The L_v measure is used to perform the model selection of mixture SEMs. Simulation studies and real data analyses are conducted to demonstrate the proposed methodologies in these chapters. Besides, to address the performances of different model selection methods, Bayes factor and DIC are also computed for model selection in this thesis. Conclusions and further developments are presented in Chapter 6, and technical details are given in the Appendix.

Chapter 2

L_v Measure for Nonlinear Structural Equation Models

2.1 Introduction

Model selection is an important issue in data analysis. Recently, many methods for model assessment and model selection have been developed. However, for structural equation models (SEMs), it is a difficult problem due to the complexity of SEMs. To deal with the problem, Bayes factor was proposed for model selection in structural equation modeling (see Jedidi et al., 1997; Lee and Song, 2001, 2003b; Lee, 2007). But as pointed out by Ibrahim et al. (2001) and Kass and Raftery (1995), this method relies on posterior model probabilities, and proper prior distributions of unknown parameters are needed. Therefore, it is usually a major task to specify the prior distributions for all models under consideration, in particular when the model space is large. Moreover, Bayes factor is generally sensitive to the choice of prior distributions, and its computational burden is heavy. Alternatively, criterion-based methods are attractive in the sense that they do not require proper prior distributions in general, and the computational burden is much light compared with Bayes factor. There are many criterion-based methods for model selection, such as Akaike information criterion (AIC) (Akaike, 1973, 1981), Bayesian information criterion (BIC) (Schwarz,

1978), and deviance information criterion (DIC) (Spiegelhalter et al., 2002). AIC and BIC are statistics for model assessment and selection based on maximum likelihood estimates. In this chapter, a statistic called the L_v measure (see Gelfand and Ghosh, 1998; Ibrahim et al., 2001; Chen et al., 2004) will be applied to model selection in nonlinear SEMs. As discussed in Ibrahim et al. (2001), the L_v measure can be written as a sum of two components, one is related to the reliability of the prediction, and the other is related to the discrepancy between the prediction and the observed data. It can be viewed as a Bayesian goodness-of-fit statistic, and can be used as a criterion for model assessment and selection. By using MCMC method, the computation of the L_v measure is quite easy after obtaining the estimates of unknown parameters and latent variables. To compare the performance of different model selection methods, Bayes factor and DIC will also be computed for model selection in this chapter.

The remainder of this chapter is divided into six sections. In Section 2.2, a brief review of the L_v measure for model selection will be given. In Section 2.3, a nonlinear SEM will be specified. In Section 2.4, the L_v measure for model selection of nonlinear SEMs will be introduced. In Section 2.5, a simulation study is presented to demonstrate the performance of the L_v measure. In Section 2.6, a real example is analyzed to illustrate the methodology. A discussion is given in Section 2.7.

2.2 Brief Review of the L_v Measure

Let $\boldsymbol{Y}^{obs} = (\boldsymbol{y}_1^{obs}, \cdots, \boldsymbol{y}_n^{obs})$ be a matrix of observations, and $\boldsymbol{Y}^{rep} = (\boldsymbol{y}_1^{rep}, \cdots, \boldsymbol{y}_n^{rep})$, which has the same distribution with $\boldsymbol{Y}^{obs}$, be the future value of an imagined replicate experiment. Suppose that for the observations in $\boldsymbol{Y}^{obs}$, a class of models denoted by $\{M_t, t = 0, 1, \cdots, T\}$ are considered. Under a certain model M_t, let $\boldsymbol{\theta}$ be the parameter vector that contains all unknown parameters in the model, and $\boldsymbol{a} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ be an estimate trying to

accommodate both $\boldsymbol{Y}^{obs}$ and $\boldsymbol{Y}^{rep}$. Then a minimum posterior predictive loss for this model was proposed (see Gelfand and Ghosh, 1998):

$$\begin{aligned} L_k(\boldsymbol{Y}^{obs}, M_t) &= \sum_{i=1}^{n} \min_{\boldsymbol{a}_i} E_{\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs},M_t} L(\boldsymbol{y}_i^{rep}, \boldsymbol{a}_i; \boldsymbol{y}_i^{obs}) \\ &= \sum_{i=1}^{n} \min_{\boldsymbol{a}_i} \left\{ E_{\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs},M_t} L(\boldsymbol{y}_i^{rep}, \boldsymbol{a}_i) + kL(\boldsymbol{y}_i^{obs}, \boldsymbol{a}_i) \right\}, \end{aligned} \tag{2.1}$$

where k is a weight that indicates the trade-off between the departure from $\boldsymbol{y}_i^{obs}$ and the departure from $\boldsymbol{y}_i^{rep}$, $L(\cdot, \cdot)$ denotes a certain loss, and different loss will give different criterion for model choice. $L(\boldsymbol{y}_i^{rep}, \boldsymbol{a}_i; \boldsymbol{y}_i^{obs})$ can be interpreted as the loss for guessing $\boldsymbol{a}_i$ when $\boldsymbol{y}_i^{rep}$ is obtained and $\boldsymbol{y}_i^{obs}$ is observed. In equation (2.1), by using the Euclidean distance defined in Ibrahim and Laud (1994), the L_v measure (see Ibrahim et al., 2001) for model M_t is defined as follows:

$$L_v(\boldsymbol{Y}^{obs}, M_t) = \sum_{i=1}^{n} \text{tr}[Var(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M_t)] + v \sum_{i=1}^{n} (\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs})^T (\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs}), \tag{2.2}$$

where $\boldsymbol{\mu}_i = E(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M_t)$, and $v = \frac{k}{k+1}$, $0 \leq v < 1$. From equation (2.1), $k \in [0, \infty)$ is a trade-off between two losses. $k = 1$ means equal weights, which makes $v = 0.5$. Therefore, in this chapter, we will consider the L_v measure with v equals to 0.5. The conditional variance and expectation in equation (2.2) are taken with respect to the posterior predictive distribution of $(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M_t)$, which is defined by

$$p(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M_t) = \int p(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, M_t) p(\boldsymbol{\theta}|\boldsymbol{Y}^{obs}, M_t) d\boldsymbol{\theta}.$$

From its definition, the L_v measure can be viewed as a Bayesian goodness-of-fit statistic, which measures the performance of a model by a combination of how close its predictions are to the observed data and the variability of the predictions. The model with the smallest value of the L_v measure will be selected. Specifically, let $P = \sum_{i=1}^{n} tr[Var(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M_t)]$ and $G = \sum_{i=1}^{n} (\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs})^T (\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs})$, then $L_v(\boldsymbol{Y}^{obs}, M_t) = P + \nu \times G$,

where P can be viewed as a penalty term, and G is an error sum of squares and can be viewed as a goodness-of-fit measure. For over-fitted model, P will decrease while G will increase; for underestimated model, P will increase while G will decrease. Therefore, complexity is penalized and a parsimonious choice is encouraged.

2.3 Model Description

Let $\boldsymbol{y}_i$, for $i = 1, \cdots, n$, be a $p \times 1$ random vector of observed variables, and $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n)$. The nonlinear SEM denoted by M is defined by

$$M: \quad \boldsymbol{y}_i = \boldsymbol{u} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \tag{2.3}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i, \tag{2.4}$$

where $\boldsymbol{u}$ is a $p \times 1$ mean vector; $\boldsymbol{\omega}_i$ is a $q \times 1$ vector of latent variables; $\boldsymbol{\epsilon}_i$ is a $p \times 1$ random vector of error terms, and is independent of $\boldsymbol{\omega}_i$; $\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$ is a partition of $\boldsymbol{\omega}_i$ into endogenous and exogenous latent vectors $\boldsymbol{\eta}_i(q_1 \times 1)$ and $\boldsymbol{\xi}_i(q_2 \times 1)$, respectively. $\boldsymbol{\Pi}$ and $\boldsymbol{\Gamma}$ are matrices of unknown regression coefficients; $\boldsymbol{F}(\cdot) = (f_1(\cdot), \cdots, f_r(\cdot))^T$ is a $r \times 1$ vector-valued function with differentiable functions $f_1, \cdots, f_r$, and $r \geq q_2$; $\boldsymbol{\delta}_i$ is a $q_1 \times 1$ random vector of error terms, and is independent of $\boldsymbol{\xi}_i$. We assume that, for $i = 1, \cdots, n$,

$$\boldsymbol{\epsilon}_i \sim N[\boldsymbol{0}, \boldsymbol{\Psi}_\epsilon], \ \boldsymbol{\xi}_i \sim N[\boldsymbol{0}, \boldsymbol{\Phi}], \ \boldsymbol{\delta}_i \sim N[\boldsymbol{0}, \boldsymbol{\Psi}_\delta]. \tag{2.5}$$

where $\boldsymbol{\Psi}_\epsilon = diag(\psi_{\epsilon 1}, \cdots, \psi_{\epsilon p})$ and $\boldsymbol{\Psi}_\delta = diag(\psi_{\delta 1}, \cdots, \psi_{\delta q_1})$.

Let $\boldsymbol{G}(\boldsymbol{\omega}_i) = (\boldsymbol{\eta}_i^T, \boldsymbol{F}(\boldsymbol{\xi}_i)^T)^T$, and $\boldsymbol{\Lambda}_\omega = (\boldsymbol{\Pi}, \boldsymbol{\Gamma})$, then equation (2.4) can be rewritten as $\boldsymbol{\eta}_i = \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{\omega}_i) + \boldsymbol{\delta}_i$. Moreover, let $\boldsymbol{\Lambda}_\eta$ and $\boldsymbol{\Lambda}_\xi$ be the submatrices of $\boldsymbol{\Lambda}$ corresponding to $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$, respectively. And let $\boldsymbol{\Pi}_0 = \boldsymbol{I}_{q_1} - \boldsymbol{\Pi}$, which is assumed to be nonsingular, then model M can be written as

$$\boldsymbol{y}_i = \boldsymbol{u} + \boldsymbol{\Lambda}_\eta \boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i) + \boldsymbol{\Lambda}_\xi \boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_i. \tag{2.6}$$

The Bayesian approach (see Dunson, 2000; Lee and Song, 2004; Lee, 2007) can be applied to the estimation of this nonlinear SEM. Here, we will

focus on model selection based on the L_v measure. For convenience, the following notations will be used. Let $\boldsymbol{Y}^{obs} = (\boldsymbol{y}_1^{obs}, \cdots, \boldsymbol{y}_n^{obs})$ be the observed continuous data set, where $\boldsymbol{y}_i^{obs} = (y_{i1}^{obs}, \cdots, y_{ip}^{obs})^T (i = 1, \cdots, n)$ is the ith column of $\boldsymbol{Y}^{obs}$, and $\boldsymbol{Y}^{rep} = (\boldsymbol{y}_1^{rep}, \cdots, \boldsymbol{y}_n^{rep})$ be the replicated data set which has the same distribution with $\boldsymbol{Y}^{obs}$, where $\boldsymbol{y}_i^{rep} = (y_{i1}^{rep}, \cdots, y_{ip}^{rep})^T$. Let $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ be the matrix of latent variables. Moreover, let $\boldsymbol{\Omega}_1 = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_n)$ and $\boldsymbol{\Omega}_2 = (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n)$ be the submatrices of $\boldsymbol{\Omega}$ corresponding to $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$, respectively. Furthermore, let $\boldsymbol{G} = (\boldsymbol{G}(\boldsymbol{\omega}_1), \cdots, \boldsymbol{G}(\boldsymbol{\omega}_n))$, and $\boldsymbol{\theta}$ be the vector that contains all the unknown elements in $\boldsymbol{u}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}_\varepsilon, \boldsymbol{\Pi}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}$ and $\boldsymbol{\Psi}_\delta$ in the model defined by equations (2.3) and (2.4). Finally, let Θ be the space of the parameter vector $\boldsymbol{\theta}$, and Ξ be the space of the latent variables $\boldsymbol{\xi}_i$, for $i = 1, \cdots, n$.

2.4 L_v Measure for Nonlinear Structural Equation Models

2.4.1 Definition of the L_v measure

In this part, the L_v measure will be applied to the nonlinear SEM, M, defined by equations (2.3) and (2.4). According to equation (2.2), with observations $\boldsymbol{Y}^{obs}$, the L_v measure for model M can be defined by

$$L_v(\boldsymbol{Y}^{obs}, M) = \sum_{i=1}^{n} \text{tr}\left\{\text{Var}(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M)\right\} + v \sum_{i=1}^{n} \text{tr}\left\{(\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs})(\boldsymbol{\mu}_i - \boldsymbol{y}_i^{obs})^T\right\}, \tag{2.7}$$

where $\boldsymbol{\mu}_i = E(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M)$, in which the expectation is taken with respect to the posterior predictive distribution of $(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M)$,

$$p(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) = \int_{\Theta \times \Xi} p(\boldsymbol{y}_i^{rep}|\boldsymbol{\xi}_i, \boldsymbol{\theta}, M) p(\boldsymbol{\xi}_i, \boldsymbol{\theta}|\boldsymbol{Y}^{obs}, M) d\boldsymbol{\theta} d\boldsymbol{\xi}_i. \tag{2.8}$$

It can be shown that, the L_v measure given by equation (2.7) can be rewritten as:

$$L_v(\boldsymbol{Y}^{obs}, M) = \sum_{j=1}^{p}\sum_{i=1}^{n} \mathrm{Var}(y_{ij}^{rep}|\boldsymbol{Y}^{obs}, M) + v\sum_{j=1}^{p}\sum_{i=1}^{n}(\mu_{ij} - y_{ij}^{obs})^2, \quad (2.9)$$

where $\mathrm{Var}(y_{ij}^{rep}|\boldsymbol{Y}^{obs}, M)$ is the jth diagonal element of $\mathrm{Var}(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M)$, and μ_{ij} is the jth element of $\boldsymbol{\mu}_i$. The first term in equation (2.9) is the penalty part and the second term is the bias part. The model with the smallest value of the L_v measure will be selected. The conditional expectation and conditional variance-covariance matrix required in the L_v measure can be given as follows:

$$\boldsymbol{\mu}_i = E(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) = E\left[E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)|\boldsymbol{Y}^{obs}, M\right], \quad (2.10)$$

and

$$\begin{aligned}\mathrm{Var}(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) = {} & E\left[\mathrm{Var}(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)|\boldsymbol{Y}^{obs}, M\right] \\ & + \mathrm{Var}\left[E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)|\boldsymbol{Y}^{obs}, M\right],\end{aligned} \quad (2.11)$$

where,

$$\begin{aligned}&\mathrm{Var}\left[E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)|\boldsymbol{Y}^{obs}, M\right] \\ &= E\left[E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)^T|\boldsymbol{Y}^{obs}, M\right] \\ &- E\left[E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)|\boldsymbol{Y}^{obs}, M\right] E\left[E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)|\boldsymbol{Y}^{obs}, M\right]^T.\end{aligned}$$

From the definition of the model given by equation (2.6), we can get

$$E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M) = \boldsymbol{u} + \boldsymbol{\Lambda}_\eta\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i)) + \boldsymbol{\Lambda}_\xi\boldsymbol{\xi}_i, \quad (2.12)$$

$$\mathrm{Var}(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M) = \boldsymbol{\Psi}_\varepsilon + \boldsymbol{\Lambda}_\eta\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Psi}_\delta(\boldsymbol{\Lambda}_\eta\boldsymbol{\Pi}_0^{-1})^T. \quad (2.13)$$

Note that the conditional expectation and variance given by equations (2.10) and (2.11) cannot be obtained directly, because the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\Omega}_1|\boldsymbol{Y}^{obs})$ contains intractable high dimensional integral. Therefore we can't get the closed form of the L_v measure, and Markov Chain Monte Carlo (MCMC) methods will be used to calculate the L_v measure for the nonlinear SEMs.

2.4.2 Computation of the L_v Measure

In this section, MCMC method for computing the L_v measure is discussed. According to the definition of the L_v measure, it can be easily calculated after the Bayesian estimates of unknown parameters and latent variables are obtained. In the Bayesian analysis, we will treat the latent variables in $\boldsymbol{\Omega}$ as hypothetical missing data, and augment the observed data set $\boldsymbol{Y}^{obs}$ with $\boldsymbol{\Omega}$ in the posterior analysis. A sufficiently large sample of $(\boldsymbol{\Omega}, \boldsymbol{\theta})$ from the joint posterior distribution $[\boldsymbol{\theta}, \boldsymbol{\Omega}|\boldsymbol{Y}^{obs}]$ can be generated by the following Gibbs sampler algorithm (see Geman and Geman, 1984). At the $(r+1)$th iteration with a current values of $(\boldsymbol{\Omega}^{(r)}, \boldsymbol{\theta}^{(r)})$:

Step a Generate $\boldsymbol{\Omega}^{(r+1)}$ from $p(\boldsymbol{\Omega}|\boldsymbol{\theta}^{(r)}, \boldsymbol{Y}^{obs})$,

Step b Generate $\boldsymbol{\theta}^{(r+1)}$ from $p(\boldsymbol{\theta}|\boldsymbol{\Omega}^{(r+1)}, \boldsymbol{Y}^{obs})$.

After the convergence of the MCMC algorithm, the samples of $\{(\boldsymbol{\theta}^{(r)}, \boldsymbol{\Omega}^{(r)}), r = 1, \cdots, R\}$ are generated. Then the Bayesian estimates $\hat{\boldsymbol{\omega}}_i$ and $\hat{\boldsymbol{\theta}}$ can be obtained as follows:

$$\hat{\boldsymbol{\omega}}_i = \frac{1}{R}\sum_{r=1}^{R} \boldsymbol{\omega}_i^{(r)}, \qquad \hat{\boldsymbol{\theta}} = \frac{1}{R}\sum_{r=1}^{R} \boldsymbol{\theta}^{(r)},$$

and the estimates of $\mathrm{Var}(\boldsymbol{\omega}_i|\boldsymbol{Y}^{obs})$ and $\mathrm{Var}(\boldsymbol{\theta}|\boldsymbol{Y}^{obs})$ can be obtained as

$$\widehat{\mathrm{Var}}(\boldsymbol{\omega}_i|\boldsymbol{Y}^{obs}) = \frac{1}{R-1}\sum_{r=1}^{R}(\boldsymbol{\omega}_i^{(r)} - \hat{\boldsymbol{\omega}}_i{}^{(r)})(\boldsymbol{\omega}_i^{(r)} - \hat{\boldsymbol{\omega}}_i{}^{(r)})^T,$$

$$\widehat{\mathrm{Var}}(\boldsymbol{\theta}|\boldsymbol{Y}^{obs}) = \frac{1}{R-1}\sum_{r=1}^{R}(\boldsymbol{\theta}^{(r)} - \hat{\boldsymbol{\theta}}^{(r)})(\boldsymbol{\theta}^{(r)} - \hat{\boldsymbol{\theta}}^{(r)})^T.$$

The estimate of the L_v measure for model M can be given as

$$\widehat{L_v}(\boldsymbol{Y}^{obs}, M) = \sum_{i=1}^{n} \mathrm{tr}(\hat{\boldsymbol{\Sigma}}) + v\sum_{i=1}^{n} \mathrm{tr}\left[\left(\hat{\boldsymbol{\mu}}_i - \boldsymbol{y}_i^{obs}\right)\left(\hat{\boldsymbol{\mu}}_i - \boldsymbol{y}_i^{obs}\right)^T\right],$$

where

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{R}\sum_{r=1}^{R}\left[\boldsymbol{u}^{(r)} + \boldsymbol{\Lambda}_\eta^{(r)}(\boldsymbol{\Pi}_0^{(r)})^{-1}(\boldsymbol{\Gamma}^{(r)}\boldsymbol{F}(\boldsymbol{\xi}_i^{(r)})) + \boldsymbol{\Lambda}_\xi^{(r)}\boldsymbol{\xi}_i^{(r)}\right],$$

and

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}} &= \frac{1}{R}\sum_{r=1}^{R}\left[\boldsymbol{\Psi}_{\varepsilon}^{(r)} + \boldsymbol{\Lambda}_{\eta}^{(r)}(\boldsymbol{\Pi}_0^{(r)})^{-1}\boldsymbol{\Psi}_{\delta}^{(r)}(\boldsymbol{\Lambda}_{\eta}^{(r)}(\boldsymbol{\Pi}_0^{(r)})^{-1})^T\right] \\ &+ \frac{1}{R}\sum_{r=1}^{R}(\boldsymbol{\mu}_i^{(r)})(\boldsymbol{\mu}_i^{(r)})^T - \left(\frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\mu}_i^{(r)}\right)\left(\frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\mu}_i^{(r)}\right)^T,\end{aligned}$$

where $\boldsymbol{\mu}_i^{(r)} = \boldsymbol{u}^{(r)} + \boldsymbol{\Lambda}_{\eta}^{(r)}(\boldsymbol{\Pi}_0^{(r)})^{-1}(\boldsymbol{\Gamma}^{(r)}\boldsymbol{F}(\boldsymbol{\xi}_i^{(r)})) + \boldsymbol{\Lambda}_{\xi}^{(r)}\boldsymbol{\xi}_i^{(r)}$.

In the Gibbs sampler, the conditional distributions $p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \boldsymbol{Y}^{obs}, M)$ and $p(\boldsymbol{\theta}|\boldsymbol{\Omega}, \boldsymbol{Y}^{obs}, M)$ are required. To obtain these conditional distributions, the prior distributions of the unknown parameters are needed to be specified. We will discuss them in the next subsection.

2.4.3 Full Conditional Distributions

In this section, the conditional distributions required in the Gibbs sampler will be discussed. Let $\boldsymbol{\theta}_y$ be the unknown parameters in $\boldsymbol{u}, \boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}_{\epsilon}$ that are associated with the measurement equation; and $\boldsymbol{\theta}_{\omega}$ be the unknown parameters in $\boldsymbol{\Gamma}, \boldsymbol{\Pi}, \boldsymbol{\Phi}$ and $\boldsymbol{\Psi}_{\delta}$ that are associated with the structural equation. For simplicity, we assume no fixed parameters. It is natural to assume that the prior distribution of $\boldsymbol{\theta}_y$ is independent of the prior distribution of $\boldsymbol{\theta}_{\omega}$, i.e. $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_y)p(\boldsymbol{\theta}_{\omega})$. Moreover, as $p(\boldsymbol{Y}, \boldsymbol{\Omega}|\boldsymbol{\theta}) = p(\boldsymbol{Y}|\boldsymbol{\Omega}, \boldsymbol{\theta}_y)p(\boldsymbol{\Omega}|\boldsymbol{\theta}_{\omega})$, we have

$$\begin{aligned}p(\boldsymbol{\theta}_y, \boldsymbol{\theta}_{\omega}|\boldsymbol{Y}, \boldsymbol{\Omega}) &\propto p(\boldsymbol{Y}, \boldsymbol{\Omega}|\boldsymbol{\theta}_y, \boldsymbol{\theta}_{\omega})p(\boldsymbol{\theta}_y)p(\boldsymbol{\theta}_{\omega}) \\ &= \left[p(\boldsymbol{Y}|\boldsymbol{\Omega}, \boldsymbol{\theta}_y)p(\boldsymbol{\theta}_y)\right]\left[p(\boldsymbol{\Omega}|\boldsymbol{\theta}_{\omega})p(\boldsymbol{\theta}_{\omega})\right].\end{aligned}$$

As a result, the conditional distributions of $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_{\omega}$ can be treated separately. The following commonly used conjugate type prior distributions are used: for $k = 1, \cdots, p$,

$$\begin{aligned}&p(\boldsymbol{u}) \triangleq N[\boldsymbol{u}_0, \boldsymbol{\Sigma}_0], \quad p(\psi_{\epsilon k}^{-1}) \triangleq Gamma[\alpha_{0\epsilon k}, \beta_{0\epsilon k}], \\ &p(\boldsymbol{\Lambda}_k|\psi_{\epsilon k}) \triangleq N[\boldsymbol{\Lambda}_{0k}, \psi_{\epsilon k}\boldsymbol{H}_{0yk}],\end{aligned} \tag{2.14}$$

where $\boldsymbol{\Lambda}_k$ is the kth row of $\boldsymbol{\Lambda}$; $\alpha_{0\epsilon k}, \beta_{0\epsilon k}, \boldsymbol{u}_0, \boldsymbol{\Lambda}_{0k}, \boldsymbol{H}_{0yk}$ and $\boldsymbol{\Sigma}_0$ are hyperparameters whose values are assumed to be given. For $k \neq h$, it is assumed that $(\psi_{\epsilon k}, \boldsymbol{\Lambda}_k)$ and $(\psi_{\epsilon h}, \boldsymbol{\Lambda}_h)$ are independent. The conditional distribution of $\boldsymbol{\Lambda}$ is given for the case where all its elements are unknown parameters. The case with fixed elements in $\boldsymbol{\Lambda}$ can be handled with slight modification.

Let $\boldsymbol{A}_k = (\boldsymbol{H}_{0yk}^{-1} + \boldsymbol{\Omega}\boldsymbol{\Omega}^T)^{-1}$, $\boldsymbol{Y}_k^* = (y_{k1}^*, \cdots, y_{kn}^*)^T$ with $y_{ki}^* = y_{ki} - u_k$, $\boldsymbol{a}_k = \boldsymbol{A}_k[\boldsymbol{H}_{0yk}^{-1}\boldsymbol{\Lambda}_{0k} + \boldsymbol{\Omega}\boldsymbol{Y}_k^*]$, and $\beta_{\epsilon k} = \beta_{0\epsilon k} + \frac{1}{2}(\boldsymbol{Y}_k^{*T}\boldsymbol{Y}_k^* - \boldsymbol{a}_k^T\boldsymbol{A}_k^{-1}\boldsymbol{a}_k + \boldsymbol{\Lambda}_{0k}^T\boldsymbol{H}_{0yk}^{-1}\boldsymbol{\Lambda}_{0k})$. Then, it can be shown that for $k = 1, \cdots, p$,

$$p(\psi_{\epsilon k}^{-1}|\boldsymbol{Y}, \boldsymbol{\Omega}, \boldsymbol{u}) \triangleq Gamma[n/2 + \alpha_{0\epsilon k}, \beta_{\epsilon k}],$$

$$p(\boldsymbol{\Lambda}_k|\boldsymbol{Y}, \boldsymbol{\Omega}, \psi_{\epsilon k}^{-1}, \boldsymbol{u}) \triangleq N[\boldsymbol{a}_k, \psi_{\epsilon k}\boldsymbol{A}_k],$$

$$p(\boldsymbol{u}|\boldsymbol{Y}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon) \triangleq N[(\boldsymbol{\Sigma}_0^{-1} + n\boldsymbol{\Psi}_\epsilon^{-1})^{-1}(n\boldsymbol{\Psi}_\epsilon^{-1}\bar{\boldsymbol{Y}} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{u}_0), (\boldsymbol{\Sigma}_0^{-1} + n\boldsymbol{\Psi}_\epsilon^{-1})^{-1}],$$

where $\bar{\boldsymbol{Y}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)$.

Now, consider the conditional distribution of $\boldsymbol{\theta}_\omega$ that is proportional to $p(\boldsymbol{\Omega}|\boldsymbol{\theta}_\omega)p(\boldsymbol{\theta}_\omega)$. Since the distribution of $\boldsymbol{\xi}$ only involves $\boldsymbol{\Phi}$, $p(\boldsymbol{\Omega}_2|\boldsymbol{\theta}_\omega) = p(\boldsymbol{\Omega}_2|\boldsymbol{\Phi})$. Moreover, it is assumed that the distribution of $\boldsymbol{\Phi}$ is independent of the prior distribution of $(\boldsymbol{\Lambda}_\omega, \boldsymbol{\Psi}_\delta)$, then we get

$$p(\boldsymbol{\Omega}|\boldsymbol{\theta}_\omega)p(\boldsymbol{\theta}_\omega) \propto [p(\boldsymbol{\Omega}_1|\boldsymbol{\Omega}_2, \boldsymbol{\Lambda}_\omega, \boldsymbol{\Psi}_\delta)p(\boldsymbol{\Lambda}_\omega, \boldsymbol{\Psi}_\delta)]\,[p(\boldsymbol{\Omega}_2|\boldsymbol{\Phi})p(\boldsymbol{\Phi})]\,.$$

Thus, the conditional distributions of $(\boldsymbol{\Lambda}_\omega, \boldsymbol{\Psi}_\delta)$ and $\boldsymbol{\Phi}$ can be treated separately.

Similar to the literature of Bayesian analysis of SEMs, the following conjugate prior distributions for the parameters in $\boldsymbol{\theta}_\omega$ are considered:

$$\begin{aligned} &p(\boldsymbol{\Phi}^{-1}) \triangleq W_{q_2}[\boldsymbol{R}_0, \rho_0], \quad p(\psi_{\delta k}^{-1}) \triangleq Gamma[\alpha_{0\delta k}, \beta_{0\delta k}], \\ &p(\boldsymbol{\Lambda}_{\omega k}|\psi_{\delta k}) \triangleq N[\boldsymbol{\Lambda}_{0\omega k}, \psi_{\delta k}\boldsymbol{H}_{0\omega k}], \end{aligned} \tag{2.15}$$

where $\boldsymbol{\Lambda}_{\omega k}$ is the kth row of $\boldsymbol{\Lambda}_\omega$; $\boldsymbol{R}_0, \rho_0, \alpha_{0\delta k}, \beta_{0\delta k}$ and $\boldsymbol{H}_{0\omega k}$ are all given hyperparameters. For $h \neq k$, $(\psi_{\delta h}, \boldsymbol{\Lambda}_{\omega h})$ and $(\psi_{\delta k}, \boldsymbol{\Lambda}_{\omega k})$ are assumed to be independent. Similar to $\boldsymbol{\Lambda}_k$, we assume no fixed elements in $\boldsymbol{\Lambda}_{\omega k}$. It can be shown that the conditional distribution of $\boldsymbol{\Phi}$ is

$$p(\boldsymbol{\Phi}|\boldsymbol{\Omega}_2) \triangleq IW_{q_2}[(\boldsymbol{\Omega}_2\boldsymbol{\Omega}_2^T + \boldsymbol{R}_0^{-1}), n + \rho_0]. \tag{2.16}$$

Let $\boldsymbol{\Omega}_{1k} = (\eta_{1k}, \cdots, \eta_{nk})^T$ for $k = 1, \cdots, q_1$. It can be shown that,

$$\begin{aligned} p(\psi_{\delta k}^{-1}|\boldsymbol{\Omega}) &\triangleq Gamma[n/2 + \alpha_{0\delta k}, \beta_{0\delta k}], \\ p(\boldsymbol{\Lambda}_{\omega k}|\boldsymbol{\Omega}, \psi_{\delta k}^{-1}) &\triangleq N[\boldsymbol{a}_{\omega k}, \psi_{\delta k}\boldsymbol{A}_{\omega k}], \end{aligned} \tag{2.17}$$

where $\boldsymbol{A}_{\omega k} = (\boldsymbol{H}_{0\omega k}^{-1} + \boldsymbol{G}\boldsymbol{G}^T)^{-1}$, $\boldsymbol{a}_{\omega k} = \boldsymbol{A}_{\omega k}[\boldsymbol{H}_{0\omega k}^{-1}\boldsymbol{\Lambda}_{0\omega k} + \boldsymbol{G}\boldsymbol{\Omega}_{1k}]$, and $\beta_{\delta k} = \beta_{0\delta k} + 2^{-1}(\boldsymbol{\Omega}_{1k}\boldsymbol{\Omega}_{1k} - \boldsymbol{a}_{\omega k}^T\boldsymbol{A}_{\omega k}^{-1}\boldsymbol{a}_{\omega k} + \boldsymbol{\Lambda}_{0\omega k}^T\boldsymbol{H}_{0\omega k}^{-1}\boldsymbol{\Lambda}_{0\omega k})$. Therefore, the conditional distributions associated with **Step b** are obtained, and they are the familiar Gamma, normal, and inverted Wishart distributions. Generate observations from these distributions are straightforward and fast.

Finally, consider the conditional distribution required in **Step a** in the Gibbs sampler. It can be shown on the basis of the definition and assumptions that

$$p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \boldsymbol{Y}) = \prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\boldsymbol{y}_i, \boldsymbol{\theta}) \propto \prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\omega}_i, \boldsymbol{\theta}) p(\boldsymbol{\eta}_i|\boldsymbol{\xi}_i, \boldsymbol{\theta}) p(\boldsymbol{\xi}_i|\boldsymbol{\theta}).$$

As $\boldsymbol{\omega}_i$ and $\boldsymbol{y}_i$ are both mutually independent, $p(\boldsymbol{\omega}_i|\boldsymbol{y}_i, \boldsymbol{\theta})$ is proportional to

$$\begin{aligned} &\exp\left\{ -\frac{1}{2}\boldsymbol{\xi}_i^T\boldsymbol{\Phi}^{-1}\boldsymbol{\xi}_i - \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T\boldsymbol{\Psi}_\epsilon^{-1}(\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) \right. \\ &\left. - \frac{1}{2}[\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{\omega}_i)]^T\boldsymbol{\Psi}_\delta^{-1}[\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{\omega}_i)] \right\}. \end{aligned} \tag{2.18}$$

Therefore the conditional distribution required in **Step a** is achieved. However, this distribution is nonstandard and complex, the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) algorithm is used to generate observations from the target density $p(\boldsymbol{\omega}_i|\boldsymbol{y}_i, \boldsymbol{\theta})$ as given in (2.18). In this algorithm, we choose $N[\boldsymbol{0}, \sigma^2\boldsymbol{\Sigma}_\omega]$ as a proposal distribution, where $\boldsymbol{\Sigma}_\omega^{-1} = \boldsymbol{\Sigma}_\delta^{-1} + \boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\omega^{-1}\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}_\delta^{-1}$ is given by

$$\boldsymbol{\Sigma}_\delta^{-1} = \begin{bmatrix} \boldsymbol{\Pi}_0^T\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Pi}_0, & -\boldsymbol{\Pi}_0^T\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Gamma}\boldsymbol{\Delta} \\ -(\boldsymbol{\Pi}_0^T\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Gamma}\boldsymbol{\Delta})^T & \boldsymbol{\Phi}^{-1} + \boldsymbol{\Delta}^T\boldsymbol{\Gamma}^T\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Gamma}\boldsymbol{\Delta} \end{bmatrix}, \tag{2.19}$$

and $\boldsymbol{\Delta} = \partial\boldsymbol{F}(\xi_i)/\partial\boldsymbol{\xi}_i^T|\boldsymbol{\xi}_i = 0$. Let $p(\cdot|\boldsymbol{\omega}, \sigma^2\boldsymbol{\Sigma}_\omega)$ be the proposal density corresponding to $N[\boldsymbol{\omega}, \sigma^2\boldsymbol{\Sigma}_\omega]$, the MH algorithm is implemented as follows: At the r-th iteration with a current value $\boldsymbol{\omega}_i^{(r)}$, a new candidate $\boldsymbol{\omega}_i^{(r+1)}$ is

generated from $p(\cdot|\boldsymbol{\omega}_i^{(r)}, \sigma^2\boldsymbol{\Sigma}_\omega)$, and accepting this new candidate with the probability

$$\min\left\{1, \frac{p(\boldsymbol{\omega}_i|\boldsymbol{y}_i, \boldsymbol{\theta})}{p(\boldsymbol{\omega}_i^{(r)}|\boldsymbol{y}_i, \boldsymbol{\theta})}\right\}. \tag{2.20}$$

The variance σ^2 is chosen such that the acceptance rate is approximately 0.25 or more (Robert et al., 1997).

2.4.4 Calibration Distribution

As pointed out by Ibrahim et al. (2001), criterion-based methods typically rely on the minimum criterion value as the basis for model selection. However, this basis is not satisfactory in general, since it does not allow a formal comparison of criterion values between two or more competing models. Thus, one of the crucial steps in using criterion-based method for model assessment and model choice is to define a calibration for the criterion. Let $L_v(\boldsymbol{Y}^{obs}, M_c)$ denote the L_v measure of the candidate model M_c, and $L_v(\boldsymbol{Y}^{obs}, M_t)$ denote the L_v measure of the true model M_t. Then given v, the difference of the L_v measures between the candidate model M_c and the true model M_t is defined as

$$D_v(\boldsymbol{Y}^{obs}, M_c) \equiv L_v(\boldsymbol{Y}^{obs}, M_c) - L_v(\boldsymbol{Y}^{obs}, M_t). \tag{2.21}$$

To calibrate the L_v measure, we need to construct the marginal distribution of $D_v(\boldsymbol{Y}^{obs}, M_c)$, denoted by $p(D_v(\boldsymbol{Y}^{obs}, M_c))$, computed with respect to the prior predictive distribution of $\boldsymbol{Y}^{obs}$ under the true model M_t:

$$p_t(\boldsymbol{Y}^{obs}) = \int\int p(\boldsymbol{Y}^{obs}|\boldsymbol{\Omega}, \boldsymbol{\theta}, M_t)p(\boldsymbol{\Omega}|\boldsymbol{\theta}, M_t)p(\boldsymbol{\theta}|M_t)d\boldsymbol{\Omega}d\boldsymbol{\theta}. \tag{2.22}$$

Note that Box (1980) and Ibrahim et al. (2001) developed similar ideas in calibration using the prior predictive distribution, and Gelman et al. (1996) developed calibration measures using the posterior predictive distribution. The proposed calibration distribution of the candidate model M_c is

$$PL_c \equiv p(D_v(\boldsymbol{Y}^{obs}, M_c)) = \int p(D_v(\boldsymbol{Y}^{obs}, M_c)|\boldsymbol{Y}^{obs})p_t(\boldsymbol{Y}^{obs})d\boldsymbol{Y}^{obs}. \tag{2.23}$$

As discussed by Bayarri and Berger (1999), this definition is appealing because it avoids the potential problem of a double use of the data. After obtaining the calibration distribution PL_c, several statistical summaries can be obtained. These include highest probability density (HPD) interval, the mean $\mu_v(M_c)$, and the standard deviation $SD_v(M_c)$ of $D_v(\boldsymbol{Y}^{obs}, M_c)$. Here, HPD interval denotes the shortest credible interval that means the interval with the highest posterior density. Note that all HPD intervals presented in this chapter as well as the subsequent chapters in this thesis are computed using a Monte Carlo (MC) method developed by Chen and Shao (1999). $\mu_v(M_c)$ measures, on the average, how close the candidate model and the true model. Specifically, if the candidate model is "close" to the true model, $\mu_v(M_c)$ is close to zero; otherwise, $\mu_v(M_c)$ will be far apart zero. $SD_v(M_c)$ measures the variability of $D_v(\boldsymbol{Y}^{obs}, M_c)$. Ibrahim et al. (2001) show that PL_c is not sensitive to choices of vague proper priors, and suitable choices of informative priors can be useful in improving the precision in the estimation of PL_c. Hence, we need to specify a proper prior distribution for $\boldsymbol{\theta}$ in calculating the calibration distribution.

As discussed previously, we can get a closed form of neither the L_v measure, nor the calibration distribution in the complex nonlinear SEMs. So MCMC method will be used again to estimate the calibration distributions of the models under consideration. Specifically, for a candidate model M_c, a sample of $D_v(\boldsymbol{Y}^{obs}, M_c)$ will be generated via Gibbs sampler algorithm, then the kernel density estimation method (Silverman, 1986; Sheather and Jones, 1991; Scott, 1992) will be used to estimate the distribution of $D_v(\boldsymbol{Y}^{obs}, M_c)$. Since the true model is usually unknown in practical applications, the model with the smallest L_v measure will be considered as the true model M_t. From the definition of the calibration distribution given by equation (2.23), the specific procedure for the estimation of PL_c can be given as follows:

(a) Generate $(\tilde{\boldsymbol{\Omega}}, \tilde{\boldsymbol{\theta}})$ from the prior predictive distribution $p(\boldsymbol{\Omega}|\boldsymbol{\theta}, M_t)p(\boldsymbol{\theta}|M_t)$, where $p(\boldsymbol{\theta}|M_t)$ is the prior distribution of $\boldsymbol{\theta}$ in model M_t,

(b) Generate a pseudo observation $\tilde{\boldsymbol{Y}}$ from $p(\boldsymbol{Y}|\tilde{\boldsymbol{\Omega}}, \tilde{\boldsymbol{\theta}})$,

(c) Set $\boldsymbol{Y}^{obs} = \tilde{\boldsymbol{Y}}$, and use the method described above to obtain the estimates of $L_v(\boldsymbol{Y}^{obs}, M_c)$ and $L_v(\boldsymbol{Y}^{obs}, M_t)$, and then calculate $D_v(\boldsymbol{Y}^{obs}, M_c)$.

Repeat (a), (b), and (c) R times, we obtain the samples $\{L_v^{(r)}(\boldsymbol{Y}^{obs}, M_c), r = 1, \cdots, R\}$, $\{L_v^{(r)}(\boldsymbol{Y}^{obs}, M_t), r = 1, \cdots, R\}$, and $\{D_v^{(r)}(\boldsymbol{Y}^{obs}, M_c), r = 1, \cdots, R\}$. Based on $\{D_v^{(r)}(\boldsymbol{Y}^{obs}, M_c), r = 1, \cdots, R\}$, we can estimate the calibration distribution PL_c via the kernel density estimation method, and the summaries of PL_c can be easily obtained.

Specifically, the mean and the standard deviation of $D_v(\boldsymbol{Y}^{obs}, M_c)$ are given by

$$\mu_v(M_c) = \frac{1}{R}\sum_{r=1}^{R} D_v^{(r)}(\boldsymbol{Y}^{obs}, M_c),$$

$$SD_v(M_c) = \frac{1}{R-1}\sum_{r=1}^{R}(D_v^{(r)}(\boldsymbol{Y}^{obs}, M_c) - \mu_v(M_c))^2.$$

Following Chen and Shao (1999), let $D_{(j)}$ be the jth smallest of $\{D_v^{(r)}(\boldsymbol{Y}^{obs}, M_c), r = 1, \cdots, R\}$, the $100(1-\alpha)\%$ HPD interval is given by $(D_{(j^*)}, D_{(j^*+[R(1-\alpha)])})$, where $[R(1-\alpha)]$ denotes the integer part of $R(1-\alpha)$, and j^* is chosen so that

$$D_{(j^*+[R(1-\alpha)])} - D_{(j^*)} = \min_{1\le j\le R-[R(1-\alpha)]}(D_{(j+[R(1-\alpha)])} - D_{(j)}).$$

The density function of PL_c, denoted by $f(D_v(\boldsymbol{Y}, M_c))$, is estimated on the basis of this sample via the following kernel method. The estimator of $f(D_v(\boldsymbol{Y}, M_c))$ is given by

$$\hat{f}(D_v(\boldsymbol{Y}, M_c); h) = \frac{1}{Rh}\sum_{r=1}^{R} K\left(\frac{D_v(\boldsymbol{Y}, M_c) - D_v^{(r)}(\boldsymbol{Y}, M_c)}{h}\right),$$

where K is the kernel function which is taken to be the Gaussian function: $K(\frac{x-x_i}{h}) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x-x_i)^2}{2h^2}\}$, and h is the smooth parameter which is chosen to minimize the risk function defined by

$$Rf(f(D_v(\boldsymbol{Y}, M_c)), \hat{f}(D_v(\boldsymbol{Y}, M_c); h)) \approx \frac{1}{4}(\sigma_K^2)^2 h^4 \int (f''(x))^2 dx + \frac{\int K^2(x)dx}{Rh},$$

where $\sigma_K^2 = \int x^2 K(x)dx$, and $f''(\cdot)$ is the second derivative of $f(\cdot)$, which is unknown and can be replaced by a reference distribution function. The reference distribution function is rescaled to have variance equal to the sample variance. In this thesis, the standard normal distribution is used as reference distribution, which yields the estimate of h as $\hat{h} = 1.06\hat{\sigma}R^{-1/5}$, where $\hat{\sigma}^2$ is the sample variance, see details in Silverman (1986). The procedure for estimating the density function can be done by R software.

2.5 A Simulation Study

In this section, a simulation study will be conducted to show the performance of the L_v measure for model selection of nonlinear SEMs. To compare the performances of different model selection methods, results of model selection related to DIC and Bayes factor will also be presented.

Model Setting

The observations in $\boldsymbol{Y}^{obs} = \{\boldsymbol{y}_i, i = 1, \cdots, n\}$ are simulated from a nonlinear SEM denoted by M_0, which is given as

$$M_0 : \boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \text{ and } \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i1}\xi_{i2} + \delta_i,$$

where

$$\boldsymbol{\Lambda}^T = \begin{pmatrix} 1.0^* & \lambda_{21} & \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{52} & \lambda_{62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{83} & \lambda_{93} \end{pmatrix},$$

where the elements with asterisks are fixed. The true values of the unknown parameters in this model are given as: $\lambda_{21} = \lambda_{31} = 0.8, \lambda_{52} = \lambda_{62} = 0.7, \lambda_{83} = \lambda_{93} = 0.8$, $\boldsymbol{\mu} = (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)^T$, $\gamma_1 = 0.6, \gamma_2 = 0.6, \gamma_3 = -0.5$, $\phi_{11} = \phi_{22} = 1.0, \phi_{12} = \phi_{21} = 0.2$, $\psi_{\varepsilon 1} = \cdots = \psi_{\varepsilon 9} = 0.5$, and $\psi_\delta = 0.5$.

Four SEMs denoted by $M_1, \cdots, M_4$ are considered as competing models. The measurement equation of model M_k $(k = 1, \cdots, 4)$ is the same as that

of M_0. The structural equations for the four models are given as follows:

$$
\begin{aligned}
M_1 &: \eta_{1i} = \gamma_{1,1}\xi_{1,i1} + \gamma_{1,2}\xi_{1,i2} + \gamma_{1,3}\xi_{1,i1}^2 + \gamma_{1,4}\xi_{1,i2}^2 + \gamma_{1,5}\xi_{1,i1}\xi_{1,i2} + \delta_{1,i}, \\
M_2 &: \eta_{2,i} = \gamma_{2,1}\xi_{2,i1} + \gamma_{2,2}\xi_{2,i2} + \delta_{2,i}, \\
M_3 &: \eta_{3,i} = \gamma_{3,1}\xi_{3,i1} + \gamma_{3,2}\xi_{3,i2} + \gamma_{3,3}\xi_{3,i1}^2 + \delta_{3,i}, \\
M_4 &: \eta_{4,i} = \gamma_{4,1}\xi_{4,i1} + \gamma_{4,2}\xi_{4,i2} + \gamma_{4,3}\xi_{4,i2}^2 + \delta_{4,i}.
\end{aligned}
$$

The prior distributions given in equations (2.14) and (2.15) will be used here. To study the impact of the prior inputs of the hyperparameters, two types of prior inputs, Prior I and Prior II, are considered.

Prior I: In the prior distributions of $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}_\epsilon$ and $\boldsymbol{\Lambda}_\omega$, the means of the normal distributions are taken as the true values of the corresponding parameters, and the covariance matrices are taken as the identity matrices with corresponding dimensions. In the prior distribution of $\boldsymbol{\Phi}$, ρ_0 and $\boldsymbol{R}_0$ in the Wishart distribution are taken to be 4 and $\boldsymbol{\Phi}_0$, respectively, where $\boldsymbol{\Phi}_0$ is the matrix with true values of ϕ_{11}, ϕ_{21} and ϕ_{22}. In the prior distributions of ψ_δ and $\psi_{\epsilon k}$, the hyperparameters in the Gamma distribution are taken to be $\alpha_{0k} = \alpha_{0\delta} = 9$ and $\beta_{0k} = \beta_{0\delta} = 4$.

Prior II: The prior inputs are given by the following ad hoc values: the means of the normal distributions are taken as zero, the covariance matrices are equal to four times of the identity matrices with appropriate dimensions, $\rho_0 = 4$, and $\boldsymbol{R}_0$ is the identity matrix, $\alpha_{0k} = \alpha_{0\delta} = 4$ and $\beta_{0k} = \beta_{0\delta} = 5$.

Furthermore, to study the impact of the sample size, we consider three different sample sizes (n=150, 300, and 600) in the simulation study. For each replication, a total of $R = 2000$ observations are collected after $H = 2000$ burn-in iterations. Three methods given bellow are conducted for model selection among M_0 to M_5. The results are obtained on the basis of 100 replications.

L_v Measure for Model Selection

Under Prior I, the calibration summaries are given in Table 2.1, where

mean($L_{0.5}$) denotes the mean value of the L_v measure with $v = 0.5$, $\mu_{0.5}$, $SD_{0.5}$, and 95% HPD denote the mean, the standard deviation, and the 95% HPD interval of the calibration distribution of the corresponding model with $v = 0.5$, respectively. From Table 2.1, we see that $\mu_{0.5}(M_2), \mu_{0.5}(M_3)$, and $\mu_{0.5}(M_4)$ are substantially larger than zero under each case. In addition, the 95% HPD intervals corresponding to M_2, M_3, and M_4 do not include zero. Thus, we can conclude that M_2, M_3, and M_4 are far from the true model. However, for model M_1, $\mu_{0.5}(M_1)$ is not significantly different from zero given the large value of $SD_{0.5}(M_1)$ under each case. Furthermore, zero is included in all the 95% HPD intervals of the calibration distribution of M_1. Therefore, we can conclude that model M_1 is close to the true model M_0, and that the performances of these two models are similar. According to the parsimonious principle, the simpler model M_0 is selected. Besides, when the sample size increases, the centers of the distributions of the calibration distributions of M_2, M_3, and M_4 become further from zero. But the center of the distribution of the calibration distribution of M_1 gets closer to zero. The reason for this phenomenon is that M_0 is nested in M_1, and the estimation of M_1 gets closer to M_0 when sample size increases. The estimated the calibration distributions presented in Figure 2.1(a) to Figure 2.1(c) also agree with these conclusions.

To study the sensitivity of model selection results to the prior inputs, the same data sets are reanalyzed under Prior II. The calibration summaries under Prior II are also given in Table 2.1, and the estimated the calibration distributions are presented in Figure 2.2(a) to Figure 2.2(c). The same conclusions can be obtained from these results. Therefore, the performance of the L_v measure for model selection seems not sensitive to the considered prior inputs and sample sizes. The Bayesian estimates of the unknown parameters under M_0 with Prior I and Prior II inputs are given in Table 2.2 and Table 2.3, respectively.

Bayes Factor for Model selection

In this part, Bayes factor will also be calculated to compare the above five models. Path sampling (Gelman and Meng, 1998) will be used to compute the Bayes factor. Let $p(\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t)$ be the log-likelihood function with a continuous parameter t in $[0, 1]$, and

$$U(\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}, \boldsymbol{\theta}, t) = \frac{d}{dt} \log p(\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t). \tag{2.24}$$

Further, let $t_{(0)} = 0 < t_{(1)} < \cdots < t_{(S-1)} \leq t_{(S)} = 1$ be fixed grids in $[0, 1]$. It follows from Lee (2007, ch.5) that an estimate of logarithm B_{10} obtained via the path sampling is given by

$$\widehat{\log B_{10}} = \frac{1}{2} \sum_{s=0}^{S-1} (t_{(s+1)} - t_{(s)})(\bar{U}_{(s+1)} + \bar{U}_{(s)}), \tag{2.25}$$

where $\bar{U}_{(s)}$ is the average of $U(\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}, \boldsymbol{\theta}, t_{(s)})$ based on all simulated observations, that is,

$$\bar{U}_{(s)} = J^{-1} \sum_{j=1}^{J} U(\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}^{(j)}, \boldsymbol{\theta}^{(j)}, t_{(s)}), \tag{2.26}$$

in which $\{(\boldsymbol{\Omega}^{(j)}, \boldsymbol{\theta}^{(j)}), j = 1, \cdots, J\}$ are simulated from $p(\boldsymbol{\Omega}, \boldsymbol{\theta}|\boldsymbol{Y}^{obs}, t_{(s)})$ via MCMC algorithm.

Model M_0 and the four candidate models $M_1, \cdots, M_4$ can be linked up with the parameter t. Specifically, the linking models have the same measurement equation as M_0 but different structural equations, which are given as follows:

$$\begin{aligned}
M_{t01} &: \eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + t\gamma_3 \xi_{i1}^2 + t\gamma_4 \xi_{i2}^2 + \gamma_5 \xi_{1,i1} \xi_{1,i2} + \delta_i, \\
M_{t02} &: \eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + (1-t)\gamma_3 \xi_{i1} \xi_{i2} + \delta_i, \\
M_{t03} &: \eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + t\gamma_3 \xi_{i1}^2 + (1-t)\gamma_4 \xi_{i1} \xi_{i2} + \delta_i, \\
M_{t04} &: \eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + t\gamma_3 \xi_{i2}^2 + (1-t)\gamma_4 \xi_{i1} \xi_{i2} + \delta_i.
\end{aligned}$$

Clearly, when $t = 0, M_{t01} = M_{t02} = M_{t04} = M_0$. When $t = 1, M_{t01} = M_1, M_{t02} = M_2, M_{t03} = M_3$, and $M_{t04} = M_4$. The log-likelihood functions

corresponding to the linking models are:

$$\begin{aligned}
&\log p_{01}(\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t) \\
&= -\frac{1}{2}\Big[c^* + \sum_{i=1}^{n} \boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i + \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_\varepsilon^{-1} (\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) \\
&+ \sum_{i=1}^{n} (\eta_i - \gamma_1 \xi_{i1} - \gamma_2 \xi_{i2} - t\gamma_3 \xi_{i1}^2 - t\gamma_4 \xi_{i2}^2 - \gamma_5 \xi_{1,i1} \xi_{1,i2})^2 / \psi_\delta \Big],
\end{aligned}$$

$$\begin{aligned}
&\log p_{02}(\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t) \\
&= \frac{1}{2}\Big[c^* + \sum_{i=1}^{n} \boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i + \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_\varepsilon^{-1} (\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) \\
&+ \sum_{i=1}^{n} (\eta_i - \gamma_1 \xi_{i1} - \gamma_2 \xi_{i2} - (1-t)\gamma_3 \xi_{i1}^2)^2 / \psi_\delta \Big],
\end{aligned}$$

$$\begin{aligned}
&\log p_{03}(\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t) \\
&= -\frac{1}{2}\Big[c^* + \sum_{i=1}^{n} \boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i + \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_\varepsilon^{-1} (\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) \\
&+ \sum_{i=1}^{n} (\eta_i - \gamma_1 \xi_{i1} - \gamma_2 \xi_{i2} - t\gamma_3 \xi_{i1}^2 - (1-t)\gamma_4 \xi_{i1} \xi_{i2})^2 / \psi_\delta \Big],
\end{aligned}$$

$$\begin{aligned}
&\log p_{04}(\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t) \\
&= -\frac{1}{2}\Big[c^* + \sum_{i=1}^{n} \boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i + \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_\varepsilon^{-1} (\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) \\
&+ \sum_{i=1}^{n} (\eta_i - \gamma_1 \xi_{i1} - \gamma_2 \xi_{i2} - t\gamma_3 \xi_{i2}^2 - (1-t)\gamma_4 \xi_{i1} \xi_{i2})^2 / \psi_\delta \Big],
\end{aligned}$$

where c^* is a constant that is equal to $n\{(p+q)\log(2\pi) + \log|\boldsymbol{\Psi}_\varepsilon| + \log|\boldsymbol{\Psi}_\delta| + \log|\boldsymbol{\Phi}|\}$. The first derivatives of these functions with respect to t are equal

to:

$$\frac{d\log p_{01}(\boldsymbol{Y}^{obs},\boldsymbol{\Omega}|\boldsymbol{\theta},t)}{dt}=\sum_{i=1}^{n}\frac{\eta_i-\gamma_1\xi_{i1}-\gamma_2\xi_{i2}-t\gamma_3\xi_{i1}^2-t\gamma_4\xi_{i2}^2-\gamma_5\xi_{1,i1}\xi_{1,i2}}{\psi_\delta(\gamma_3\xi_{i1}^2+\gamma_4\xi_{i2}^2)^{-1}},$$

$$\frac{d\log p_{02}(\boldsymbol{Y}^{obs},\boldsymbol{\Omega}|\boldsymbol{\theta},t)}{dt}=-\sum_{i=1}^{n}\frac{\eta_i-\gamma_1\xi_{i1}-\gamma_2\xi_{i2}-(1-t)\gamma_3\xi_{i1}^2}{\psi_\delta(\gamma_3\xi_{i1}^2)^{-1}},$$

$$\frac{d\log p_{03}(\boldsymbol{Y}^{obs},\boldsymbol{\Omega}|\boldsymbol{\theta},t)}{dt}=\sum_{i=1}^{n}\frac{\eta_i-\gamma_1\xi_{i1}-\gamma_2\xi_{i2}-t\gamma_3\xi_{i1}^2-(1-t)\gamma_4\xi_{i1}\xi_{i2}}{\psi_\delta(\gamma_3\xi_{i1}^2-\gamma_4\xi_{i1}\xi_{i2})^{-1}},$$

$$\frac{d\log p_{04}(\boldsymbol{Y}^{obs},\boldsymbol{\Omega}|\boldsymbol{\theta},t)}{dt}=\sum_{i=1}^{n}\frac{\eta_i-\gamma_1\xi_{i1}-\gamma_2\xi_{i2}-t\gamma_3\xi_{i2}^2-(1-t)\gamma_4\xi_{i1}\xi_{i2}}{\psi_\delta(\gamma_3\xi_{i2}^2-\gamma_4\xi_{i1}\xi_{i2})^{-1}}.$$

These derivatives give $U(\boldsymbol{Y}^{obs},\boldsymbol{\Omega}|\boldsymbol{\theta},t)$ for computing the logarithm Bayes factors, see equations (2.24) to (2.26).

In the path sampling procedure, we take $S=20$ grids in $[0,1]$. Based on the previous analysis, for each $t_{(s)}$, we take a burn-in phase of 2000 iterations, and further collect 2000 observations in computing the logarithm Bayes factor. The estimated logarithm Bayes factors are presented in Table 2.4. From this table, all the logarithm Bayes factors are negative, which consistently select M_0 (see the interpretation of Bayes factor in Kass and Raftery, 1995). Furthermore, the estimated logarithm Bayes factors have similar values under Prior I and Prior II inputs. Therefore, the same conclusion is drawn under these two different prior inputs.

Model Selection Using DIC and WinBUGS

In this part, DIC values with WinBUGS for model selection among M_0 to M_5 will be presented. For nonlinear SEMs, the software WinBUGS can produce Bayeian estimates of the structural parameters and latent variables, as well as DIC values for model selection. Same as before, we take the burning iterations as 2000, and further collect 2000 samples after burn-in phase. The DIC values under Prior I inputs with different sample sizes are given in Table 2.5. From this table, the true model M_0 with smallest DIC value is selected under each given sample size.

2.6 A Real Example

To illustrate the application of the L_v measure for model selection to a real data, a small portion of the Inter-university Consortium for Political and Social Research (ICPSR) data set collected in the project WORLD VALUE SURVEY 1981-1984 and 1990-1993 (World Value Study Group, ICPSR Version) is analyzed. In this example, only the data obtained from the United Kingdom were used. Six variables in the original data set (variables 180, 96, 62, 179, 116 and 117; see the Appendix) relating to the respondents' job, religions belief, and homelife were taken as observed variables in $\boldsymbol{y} = (y_1, \cdots, y_6)^T$. After deleting cases with missing data, the sample size was 197. Among them, (y_1, y_2) were related to homelife (η), (y_3, y_4) were related to religions belief (ξ_1), and (y_5, y_6) were related to job satisfaction (ξ_2). Variable (v_3, v_4) were measured in a five-point scale, while the others were measured in a ten-point scale. As the purpose of this example is for illustration, they were all treated as continuous for brevity.

For this data set, five SEMs with latent factors $\boldsymbol{\omega} = (\eta, \xi_1, \xi_2)^T$ are considered. Specifically, they are assumed to have the same measurement equation, $\boldsymbol{y} = \boldsymbol{u} + \boldsymbol{\Lambda\omega} + \boldsymbol{\epsilon}$, and different structural equations:

$$
\begin{aligned}
M_0 &: \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \delta, \\
M_1 &: \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1^2 + \delta, \\
M_2 &: \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_2^2 + \delta, \\
M_3 &: \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \delta, \\
M_4 &: \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1^2 + \gamma_4\xi_2^2 + \gamma_5\xi_1\xi_2 + \delta.
\end{aligned}
$$

In the measurement equation,

$$
\boldsymbol{\Lambda}^T = \begin{bmatrix} 1.0^* & \lambda_{21} & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 1.0^* & \lambda_{42} & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{63} \end{bmatrix},
$$

where the zero's and one's in $\boldsymbol{\Lambda}$ were treated as fixed parameters. To

calculate the L_v measures, the following hyperparameters were specified: $\alpha_{0\varepsilon k} = \alpha_{0\delta} = 10, \beta_{0\varepsilon k} = \beta_{0\delta} = 8$, $\boldsymbol{H}_{0yk}$ and $\boldsymbol{H}_{0\omega k}$ are diagonal matrices with diagonal elements 0.25, $\rho_0 = 20, \boldsymbol{\Sigma}_0 = \boldsymbol{I}_6$, $\boldsymbol{R}_0^{-1} = 2\tilde{\boldsymbol{\Phi}}, \boldsymbol{u}_0 = \tilde{\boldsymbol{u}}$ and $\boldsymbol{\Lambda}_{0k} = \tilde{\boldsymbol{\Lambda}}_{0k}$, where $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\Phi}}$, and $\tilde{\boldsymbol{\Lambda}}_{0k}$ were the Bayesian estimates obtained using noninformative prior inputs. Based on different starting values of the parameters, three parallel sequences of observations were generated. We found that the MCMC algorithm converged after about 2000 iterations. After convergence, a total of 2000 observations were collected to obtain the results. Take $v = 0.5$, the L_v measures for the five models are:

$$
\begin{aligned}
&L_{0.5}(Y^{obs}, M_0) = 3657.8, \quad L_{0.5}(Y^{obs}, M_1) = 3652.67, \\
&L_{0.5}(Y^{obs}, M_2) = 3702.8, \quad L_{0.5}(Y^{obs}, M_3) = 3568.4, \\
&L_{0.5}(Y^{obs}, M_4) = 3853.5.
\end{aligned}
$$

Since the value of the L_v measure for model M_3 is less then the others, M_3 is selected. Now we consider the calibration distributions. First, 100 replication data sets are generated based on M_3 according to the prior distribution $p(\boldsymbol{\theta})$ under the auxiliary prior inputs. Then the calibration distributions can be obtained through the method given in section 2.4.4. The summaries of the calibration distributions are presented in Table 2.6, and the densities of them are presented in Figure 2.3.

From Table 2.6, the mean values of all the calibration distributions are larger than zero, which means that, on average, the performance of M_3 is better than the others. In addition, the 95% HPD interval of M_2 dose not include zero, we conclude that M_3 performs significantly better than M_2. For M_0, although zero is included in the 95% HPD interval, the lower bound of this interval is quite close to zero. Therefore, we conclude M_3 performs better than M_0. For M_4, zero is included in the 95% HPD interval and the lower bound of this interval is not close to zero, indicating that the improvement of adding two terms, ξ_1^2 and ξ_2^2, into the structural equation of M_3 is not significant. This phenomenon can also be revealed by the

small values of $\hat{\gamma}_3(0.032)$ and $\hat{\gamma}_4(-0.061)$ in model M_4. For model M_1, zero is also included in the HPD interval. However, compared with the higher bound, the lower bound of this interval is relatively small. Considering the small value of $\hat{\gamma}_3(0.032)$ in M_1, M_3 is selected. The estimated the calibration distributions presented in Figure 2.3 give the same conclusion. The Bayesian estimates of the unknown parameters, together with their standard errors estimates, in model M_3 are given in Table 2.7. Thus, the estimated nonlinear structural equation in the selected model is

$$\eta_i = 0.376\xi_{i1} + 0.584\xi_{i2} - 0.212\xi_{i1}\xi_{i2} + \delta_i.$$

Bayes factor and DIC are also calculated for the model selection. Results of logarithm Bayes factors are:

$$\log B_{03} = -2.678, \quad \log B_{13} = -4.3134,$$
$$\log B_{23} = -2.874, \quad \log B_{43} = -0.1231.$$

All the estimated logarithm Bayes factors are negative. According to the criterion in interpreting logarithm Bayes factors (Kass and Raftery, 1995), M_3 is selected.

DIC values from WinBUGS for the five models are:

$$DIC_0 = 4093.03, \quad DIC_1 = 4090.54, \quad DIC_2 = 4093.96,$$
$$DIC_3 = 4081.56, \quad DIC_4 = 4087.63.$$

The DIC value of M_3 is smaller than the others, therefore, M_3 is selected.

2.7 Discussion

From the numerical studies given in the previous sections, the L_v measure, Bayes factor, and DIC can achieve the same conclusion in model selection. However, the computational burden of Bayes factor is heavy. For example, when taking $S = 20$ in the path sampling in calculating Bayes factor,

the computing time is almost twenty times of that for calculating the L_v measure. When applying DIC method, we select the model only according to the minimum DIC value. However, when the difference of DIC values between two competing models is small, we can't decide which one is better. As compared with the other two methods, the computation of the L_v measure is quite simple and fast. Moreover, besides considering the model with the smallest value of the L_v measure, the corresponding calibration distribution is also used to help making decision. Therefore, the L_v measure provides better alternative method for model selection of SEMs.

Table 2.1: Mean values of the L_v measure and calibration summaries for simulation study

	Sample size	Model	mean($L_{0.5}$)	$\mu_{0.5}$	$SD_{0.5}$	95% HPD
Prior I	n=150	M_0	1337.144	-	-	-
		M_1	1347.115	9.971	12.445	(-13.154, 37.325)
		M_2	1442.170	105.027	55.201	(21.066, 223.426)
		M_3	1427.118	89.974	49.080	(9.190, 206.845)
		M_4	1424.956	87.813	45.883	(11.960, 193.298)
	n=300	M_0	2688.229	-	-	-
		M_1	2699.543	11.314	11.562	(-16.137, 28.206)
		M_2	2908.353	220.124	68.943	(109.608, 385.510)
		M_3	2867.839	179.610	57.984	(70.835, 280.530)
		M_4	2873.309	185.080	60.002	(82.646, 328.037)
	n=600	M_0	5367.910	-	-	-
		M_1	5364.678	-3.232	15.443	(-34.846, 22.448)
		M_2	5838.367	470.458	108.807	(261.086, 656.186)
		M_3	5765.362	397.453	98.704	(210.631, 567.128)
		M_4	5771.552	403.642	98.070	(212.586, 572.154)
Prior II	n=150	M_0	1506.464	-	-	-
		M_1	1500.840	-5.624	8.324	(-21.980, 8.684)
		M_2	1603.918	97.455	51.061	(8.366, 202.987)
		M_3	1580.422	73.959	39.936	(10.483, 160.038)
		M_4	1581.718	75.255	43.230	(9.947, 152.681)
	n=300	M_0	2876.646	-	-	-
		M_1	2870.890	-5.756	13.911	(-34.010, 16.984)
		M_2	3084.528	207.882	65.212	(83.901, 345.813)
		M_3	3041.908	165.262	55.203	(82.329, 279.141)
		M_4	3046.956	170.310	57.366	(68.674, 289.637)
	n=600	M_0	5567.036	-	-	-
		M_1	5554.704	-12.332	19.301	(-43.150, 29.024)
		M_2	6022.318	455.282	106.536	(264.944, 652.684)
		M_3	5944.422	377.386	95.427	(212.204, 532.871)
		M_4	5949.764	382.728	96.519	(202.196, 556.094)

Table 2.2: The Bayesian estimates for simulation study under Prior I inputs

		$n = 150$			$n = 300$			$n = 600$		
Para	TRUE	mean	RMS	BIAS	mean	RMS	BIAS	mean	RMS	BIAS
λ_{21}	0.800	0.776	0.068	0.024	0.795	0.047	0.005	0.794	0.035	0.006
λ_{31}	0.800	0.778	0.055	0.022	0.796	0.047	0.004	0.796	0.036	0.004
λ_{52}	0.700	0.660	0.082	0.040	0.671	0.068	0.029	0.686	0.046	0.014
λ_{62}	0.700	0.670	0.067	0.030	0.672	0.065	0.028	0.687	0.046	0.013
λ_{83}	0.800	0.745	0.097	0.055	0.764	0.069	0.036	0.787	0.047	0.013
λ_{93}	0.800	0.752	0.085	0.048	0.763	0.062	0.037	0.784	0.048	0.016
$\psi_{\epsilon 1}$	0.500	0.478	0.073	0.022	0.501	0.060	0.001	0.493	0.040	0.007
$\psi_{\epsilon 2}$	0.500	0.533	0.076	0.033	0.504	0.049	0.004	0.503	0.036	0.003
$\psi_{\epsilon 3}$	0.500	0.510	0.065	0.010	0.504	0.048	0.004	0.509	0.037	0.009
$\psi_{\epsilon 4}$	0.500	0.476	0.072	0.024	0.476	0.067	0.024	0.480	0.050	0.020
$\psi_{\epsilon 5}$	0.500	0.528	0.067	0.028	0.509	0.050	0.009	0.504	0.033	0.004
$\psi_{\epsilon 6}$	0.500	0.522	0.069	0.022	0.508	0.047	0.008	0.509	0.037	0.009
$\psi_{\epsilon 7}$	0.500	0.470	0.074	0.030	0.475	0.066	0.025	0.482	0.053	0.018
$\psi_{\epsilon 8}$	0.500	0.528	0.071	0.028	0.526	0.054	0.026	0.508	0.037	0.008
$\psi_{\epsilon 9}$	0.500	0.527	0.077	0.027	0.517	0.051	0.017	0.509	0.040	0.009
ψ_{δ}	0.500	0.578	0.109	0.078	0.541	0.078	0.041	0.525	0.060	0.025
ϕ_{11}	1.000	1.061	0.166	0.061	1.035	0.120	0.035	1.018	0.096	0.018
ϕ_{12}	0.200	0.198	0.105	0.002	0.205	0.075	0.005	0.195	0.055	0.005
ϕ_{22}	1.000	1.051	0.181	0.051	1.037	0.127	0.037	1.020	0.087	0.020
γ_1	0.600	0.555	0.102	0.045	0.563	0.081	0.037	0.594	0.055	0.006
γ_2	0.600	0.564	0.101	0.036	0.584	0.073	0.016	0.588	0.055	0.012
γ_3	-0.500	-0.429	0.114	0.071	-0.452	0.085	0.048	-0.470	0.055	0.030
μ_1	0.000	-0.004	0.102	0.004	0.002	0.078	0.002	-0.006	0.054	0.006
μ_2	0.000	-0.004	0.095	0.004	-0.005	0.071	0.005	-0.005	0.048	0.005
μ_3	0.000	0.002	0.079	0.002	-0.001	0.064	0.001	-0.005	0.050	0.005
μ_4	0.000	0.018	0.094	0.018	0.010	0.066	0.010	-0.003	0.046	0.003
μ_5	0.000	0.012	0.068	0.012	-0.002	0.056	0.002	-0.003	0.038	0.003
μ_6	0.000	0.010	0.077	0.010	0.006	0.058	0.006	-0.001	0.038	0.001
μ_7	0.000	0.012	0.106	0.012	0.007	0.074	0.007	0.004	0.048	0.004
μ_8	0.000	0.011	0.089	0.011	0.004	0.061	0.004	0.001	0.042	0.001
μ_9	0.000	0.014	0.082	0.014	0.007	0.062	0.007	-0.002	0.044	0.002

Table 2.3: The Bayesian estimates for simulation study under Prior II inputs

		$n = 150$			$n = 300$			$n = 600$		
Para	TRUE	mean	RMS	BIAS	mean	RMS	BIAS	mean	RMS	BIAS
λ_{21}	0.800	0.757	0.073	0.043	0.783	0.048	0.017	0.788	0.035	0.012
λ_{31}	0.800	0.758	0.062	0.042	0.784	0.049	0.016	0.790	0.036	0.010
λ_{52}	0.700	0.665	0.076	0.035	0.675	0.065	0.025	0.687	0.045	0.013
λ_{62}	0.700	0.675	0.063	0.025	0.676	0.063	0.024	0.689	0.045	0.011
λ_{83}	0.800	0.745	0.093	0.055	0.766	0.066	0.034	0.788	0.045	0.012
λ_{93}	0.800	0.753	0.081	0.047	0.765	0.060	0.035	0.785	0.048	0.015
$\psi_{\epsilon 1}$	0.500	0.542	0.081	0.042	0.530	0.066	0.030	0.508	0.040	0.008
$\psi_{\epsilon 2}$	0.500	0.593	0.117	0.093	0.535	0.061	0.035	0.519	0.041	0.019
$\psi_{\epsilon 3}$	0.500	0.570	0.096	0.070	0.535	0.060	0.035	0.525	0.044	0.025
$\psi_{\epsilon 4}$	0.500	0.565	0.096	0.065	0.524	0.065	0.024	0.503	0.044	0.003
$\psi_{\epsilon 5}$	0.500	0.583	0.103	0.083	0.536	0.061	0.036	0.517	0.038	0.017
$\psi_{\epsilon 6}$	0.500	0.577	0.104	0.077	0.535	0.059	0.035	0.523	0.042	0.023
$\psi_{\epsilon 7}$	0.500	0.549	0.084	0.049	0.517	0.062	0.017	0.504	0.050	0.004
$\psi_{\epsilon 8}$	0.500	0.586	0.109	0.086	0.555	0.074	0.055	0.523	0.042	0.023
$\psi_{\epsilon 9}$	0.500	0.583	0.113	0.083	0.545	0.067	0.045	0.523	0.047	0.023
ψ_{δ}	0.500	0.673	0.190	0.173	0.590	0.111	0.090	0.552	0.076	0.052
ϕ_{11}	1.000	1.040	0.153	0.040	1.022	0.119	0.022	1.014	0.095	0.014
ϕ_{12}	0.200	0.205	0.106	0.005	0.207	0.078	0.007	0.196	0.055	0.004
ϕ_{22}	1.000	1.043	0.173	0.043	1.028	0.119	0.028	1.016	0.087	0.016
γ_1	0.600	0.566	0.098	0.034	0.571	0.077	0.029	0.597	0.052	0.003
γ_2	0.600	0.567	0.099	0.033	0.588	0.070	0.012	0.588	0.055	0.012
γ_3	-0.500	-0.439	0.109	0.061	-0.460	0.081	0.040	-0.473	0.055	0.027
μ_1	0.000	-0.002	0.108	0.002	0.002	0.075	0.002	-0.005	0.053	0.005
μ_2	0.000	-0.004	0.100	0.004	-0.006	0.068	0.006	-0.005	0.047	0.005
μ_3	0.000	0.002	0.083	0.002	-0.003	0.061	0.003	-0.005	0.049	0.005
μ_4	0.000	0.015	0.097	0.015	0.009	0.068	0.009	-0.002	0.045	0.002
μ_5	0.000	0.010	0.070	0.010	-0.003	0.057	0.003	-0.003	0.038	0.003
μ_6	0.000	0.008	0.080	0.008	0.006	0.059	0.006	-0.000	0.037	0.000
μ_7	0.000	0.011	0.105	0.011	0.005	0.073	0.005	0.005	0.047	0.005
μ_8	0.000	0.009	0.087	0.009	0.003	0.062	0.003	0.002	0.042	0.002
μ_9	0.000	0.013	0.084	0.013	0.006	0.062	0.006	-0.001	0.044	0.001

Table 2.4: Logarithm Bayes factor for simulation study

Prior	Logarithm Bayes factor	n=150	n=300	n=600
Prior I	$\log B_{10}$	-2.34	-2.965	-3.48
	$\log B_{20}$	-11.04	-22.504	-48.358
	$\log B_{30}$	-8.89	-19.697	-43.046
	$\log B_{40}$	-11.47	-21.295	-44.794
Prior II	$\log B_{10}$	-2.39	-2.95	-3.4
	$\log B_{20}$	-9.37	-20.65	-46.37
	$\log B_{30}$	-7.82	-18.34	-41.35
	$\log B_{40}$	-9.1	-18.89	-41.96

Table 2.5: DIC for simulation study with different sample sizes

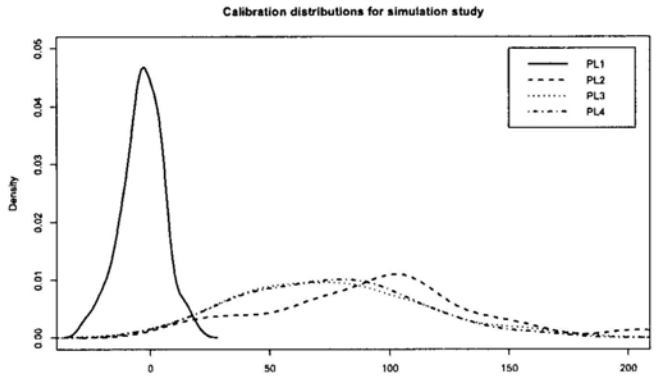
model	n=150	n=300	n=600
M_0	3233.15	6461.76	12893.5
M_1	3238.23	6465.55	12898.3
M_2	3255.77	6493.9	12961.6
M_3	3256.17	6496.7	12966.8
M_4	3255.41	6496.4	12962.2

Table 2.6: Calibration summaries for real example

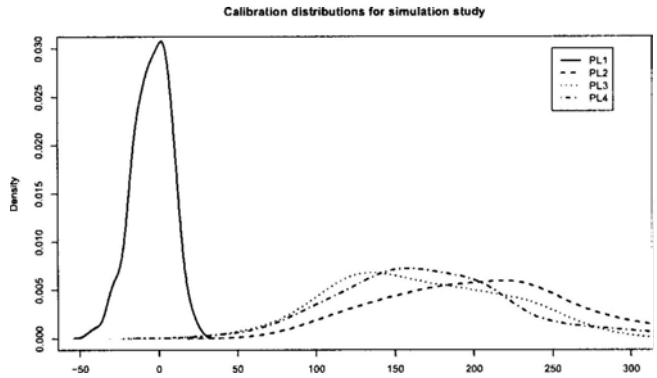
Model	$\mu_{0.5}$	$SD_{0.5}$	95% HPD
M_0	101.531	66.392	(-0.007, 259.848)
M_1	82.341	66.838	(-24.141, 203.688)
M_2	100.543	62.803	(10.458, 225.316)
M_4	21.451	27.647	(-27.615, 77.312)

Table 2.7: The Bayesian estimates for real example

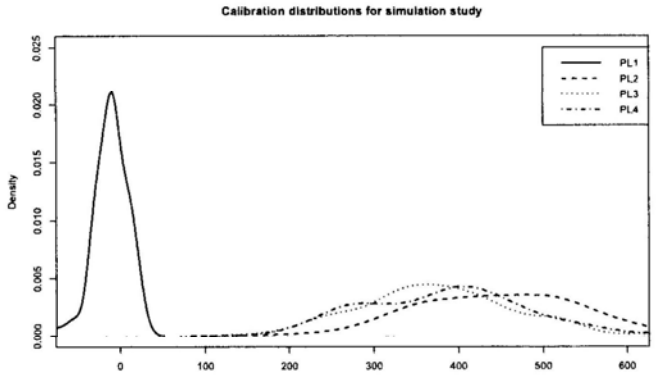
Parameter	EST	SE	Parameter	EST	SE
u_1	8.397	0.125	$\psi_{\varepsilon 1}$	0.835	0.173
u_2	7.803	0.128	$\psi_{\varepsilon 2}$	1.204	0.19
u_3	2.360	0.118	$\psi_{\varepsilon 3}$	0.873	0.15
u_4	3.047	0.126	$\psi_{\varepsilon 4}$	0.934	0.161
u_5	7.555	0.156	$\psi_{\varepsilon 5}$	1.943	0.443
u_6	7.403	0.174	$\psi_{\varepsilon 6}$	3.691	0.493
λ_{21}	0.924	0.107	γ_1	0.376	0.116
λ_{32}	-1.066	0.108	γ_2	0.584	0.104
λ_{63}	0.867	0.161	γ_3	-0.212	0.071
ϕ_{11}	1.538	0.268	ϕ_{12}	-0.209	0.2
ϕ_{22}	2.765	0.584	ψ_δ	0.846	0.177



(a) n=150

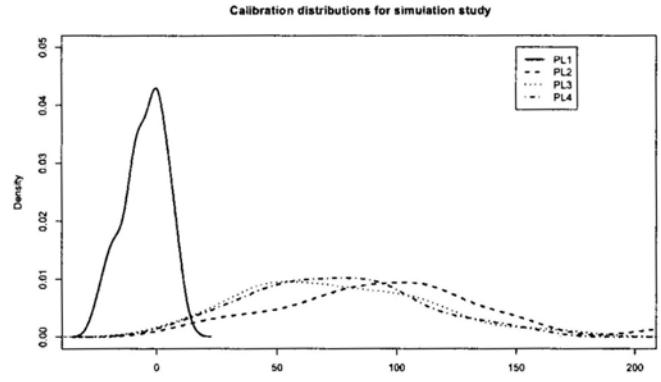


(b) n=300

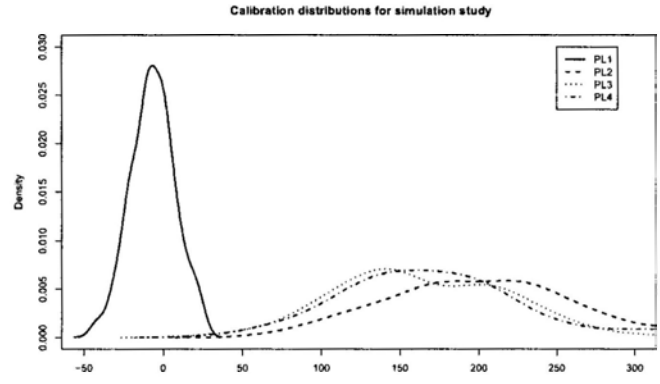


(c) n=600

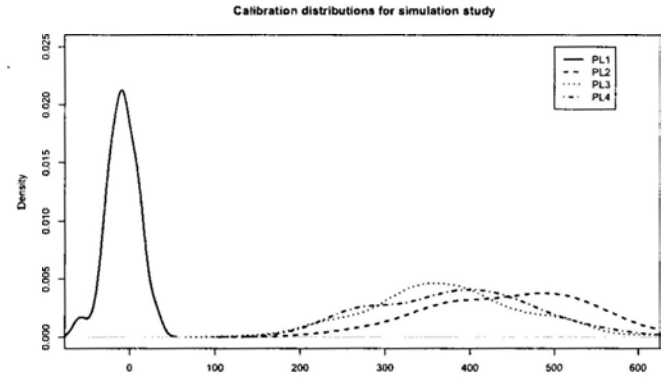
Figure 2.1: Calibration distributions under Prior I



(a) n=150



(b) n=300



(c) n=600

Figure 2.2: Calibration distributions under Prior II

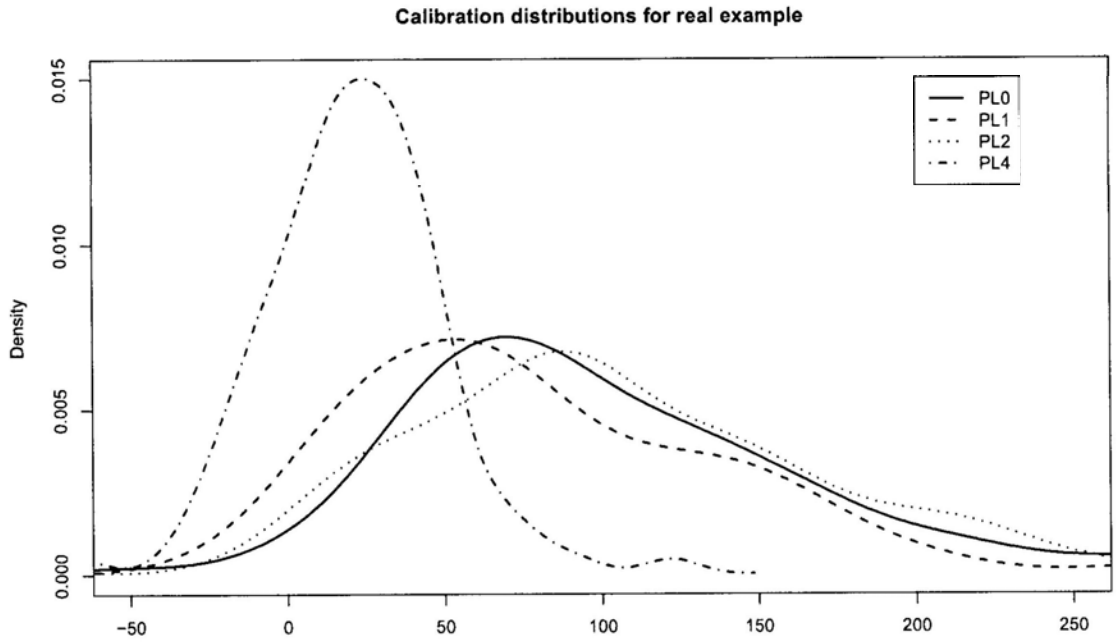


Figure 2.3: Real example

Chapter 3

L_v Measure for Nonlinear Structural Equation Models with Mixed Continuous and Categorical Data

3.1 Introduction

As discussed in the literature, structural equation models (SEMs) have been widely used in behavioral, educational, medical and social sciences. In these fields, categorical variables are often encountered. A typical example is when a subject is asked to report the opinion about a policy on scales like 'strongly disagree', 'disagree', 'no opinion', 'agree', 'strongly agree', or to report the effect of a drug on scales like 'getting worse', 'no change', 'getting better'. To deal with this kind of data, SEMs with ordered categorical variables are proposed. In the analysis of SEMs with ordered categorical data, a commonly used approach is to treat the variables as observations that come from a hidden continuous normal distribution with a threshold specification, see Lee (2007). In this chapter, the methodology developed in Chapter 2 is extended to nonlinear SEMs with mixed continuous and ordered categorical data. To define the L_v measure for categorical data, a method proposed by Chen et al. (2004) is used to transform the categorical data into binary data. The Bayesian approach, together with MCMC

algorithms, is applied to the estimation of latent variables and unknown parameters, and to the computation of the L_v measure for model selection.

The remainder of this chapter is organized as follows. Section 3.2 defines a nonlinear SEM with mixed continuous and ordered categorical data. Section 3.3 discusses the definition and computation of the L_v measure for nonlinear SEMs in the presence of ordered categorical data. In Section 3.4, a simulation study is presented to demonstrate the performance of the L_v measure in model selection of nonlinear SEMs with mixed continuous and ordered categorical data. In Section 3.5, a real example is analyzed to illustrate the methodology. A discussion is given in Section 3.6.

3.2 Model Description

Suppose $\boldsymbol{y}_i$ is a $p \times 1$ vector of manifest variables, and it satisfies the following measurement equation:

$$\boldsymbol{y}_i = \boldsymbol{u} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, i = 1, \cdots, n, \tag{3.1}$$

where $\boldsymbol{u}$ is a $p \times 1$ mean vector, $\boldsymbol{\omega}_i$ is a $q \times 1$ vector of latent variables, $\boldsymbol{\Lambda}$ is a $p \times q$ loading matrix, $\boldsymbol{\epsilon}_i$ is a vector of error measurements with distribution $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$, $\boldsymbol{\Psi}_\epsilon = \text{diag}(\psi_{\epsilon 1}, \cdots, \psi_{\epsilon p})$, and $\boldsymbol{\omega}_i$ is independent of $\boldsymbol{\epsilon}_i$. In addition, the latent vector $\boldsymbol{\omega}_i$ is partitioned into $(\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$ which satisfy the following structural equation:

$$\boldsymbol{\eta}_i = \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i, \tag{3.2}$$

where $\boldsymbol{\eta}_i$ is a $q_1 \times 1$ vector of endogenous latent variables, $\boldsymbol{\xi}_i$ is a $q_2 \times 1$ vector of exogenous latent variables, $\boldsymbol{F}(\boldsymbol{\xi}_i) = (f_1(\boldsymbol{\xi}_i), \cdots, f_m(\boldsymbol{\xi}_i))^T$ is a vector-valued function with differentiable functions $f_1, \cdots, f_m$, and $m \geq q_2$. Moreover, we assume that $\boldsymbol{\Pi}_0 = \boldsymbol{I}_{q_1} - \boldsymbol{\Pi}$ is positive definite, and its determinant is independent of the elements in $\boldsymbol{\Pi}$, $\boldsymbol{\xi}_i \sim N[\mathbf{0}, \boldsymbol{\Phi}]$ and $\boldsymbol{\delta}_i \sim N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$ are independent, and $\boldsymbol{\Psi}_\delta = \text{diag}(\psi_{\delta 1}, \cdots, \psi_{\delta q_1})$. Let $\boldsymbol{\Lambda}_\omega = (\boldsymbol{\Pi}, \boldsymbol{\Gamma})$ and $\boldsymbol{G}(\boldsymbol{\omega}_i) = (\boldsymbol{\eta}_i^T, \boldsymbol{F}(\boldsymbol{\xi}_i)^T)^T$, the structural equation given by equation (3.2) can be written as: $\boldsymbol{\eta}_i = \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{\omega}_i) + \boldsymbol{\delta}_i$.

Furthermore, let $\boldsymbol{\Lambda}_\eta$ and $\boldsymbol{\Lambda}_\xi$ are the submatrices of $\boldsymbol{\Lambda}$ corresponding to $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$, respectively. The nonlinear SEM defined by equations (3.1) and (3.2) can be rewritten as:

$$\boldsymbol{y}_i = \boldsymbol{u} + \boldsymbol{\Lambda}_\eta \boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i) + \boldsymbol{\Lambda}_\xi \boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i, i = 1, \cdots, n. \tag{3.3}$$

To deal with ordered categorical data, suppose $\boldsymbol{y}_i = (\boldsymbol{y}_{o,i}^T, \boldsymbol{y}_{u,i}^T)^T$, where $\boldsymbol{y}_{o,i}$ is a $r \times 1$ vector corresponding to the observed continuous variables, and $\boldsymbol{y}_{u,i}$ is a $s \times 1$ vector corresponding to the unobservable continuous variables. The information of $\boldsymbol{y}_{u,i}$ is given by the observed ordered categorical variables in $\boldsymbol{z}_i$. The relationship between $\boldsymbol{y}_{u,i} = (y_{u,i1}, \cdots, y_{u,is})^T$ and $\boldsymbol{z}_i$ is given as follows: for $i = 1, \cdots, n$ and $k = 1, \cdots, s$,

$$\boldsymbol{z}_i = \begin{bmatrix} z_{i1} \\ \vdots \\ z_{is} \end{bmatrix} \quad \text{if} \quad \begin{cases} \alpha_{1,z_{i1}} < y_{u,i1} \leq \alpha_{1,z_{i1}+1} \\ \vdots \\ \alpha_{s,z_{is}} < y_{u,is} \leq \alpha_{s,z_{is}+1} \end{cases}, \tag{3.4}$$

where z_{ik} is an integer that belongs to the set $\{0, 1, \cdots, b_k\}$, and $\boldsymbol{\alpha}_k = \{\alpha_{k,0}, \cdots, \alpha_{k,b_k+1}\}$ are thresholds. In general, we set $\alpha_{k,0} = -\infty, \alpha_{k,b_k+1} = \infty$. Thus, for the kth variable, there are $b_k + 1$ categories which are defined by the unknown thresholds $\boldsymbol{\alpha}_k$. For convenience, we denote this model by M.

As discussed in Lee and Song (2004), without imposing any identification conditions, models with ordered categorical variables are unidentified. So, we fix $\alpha_{k,1}$ and $\alpha_{k,b_k}(k = 1, \cdots, s)$ at some preassigned values. To identify the covariance structure of $\boldsymbol{y}_i$, following the common practice, we fix some appropriate elements in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Pi}$ at preassigned values. In the following part of this chapter, some notations are used. Let $\boldsymbol{Y}^{obs} = (\boldsymbol{Y}_o^{obs}, \boldsymbol{Y}_u)$, where $\boldsymbol{Y}_o^{obs} = (\boldsymbol{y}_{o,1}^{obs}, \cdots, \boldsymbol{y}_{o,n}^{obs})$ be the matrix of observations corresponding to the observed continuous variables, $\boldsymbol{Y}_u = (\boldsymbol{y}_{u,1}, \cdots, \boldsymbol{y}_{u,n})$ be the matrix of observations related to the underlying unobserved continuous variables, $\boldsymbol{y}_{o,i}^{obs} = (y_{o,i1}^{obs}, \cdots, y_{o,ir}^{obs})^T (i = 1, \cdots, n)$, $\boldsymbol{y}_{u,i} = (\boldsymbol{y}_{u,i1}, \cdots, \boldsymbol{y}_{u,is})^T$ and

$\boldsymbol{y}_i = (\boldsymbol{y}_{o,i}^{obs^T}, \boldsymbol{y}_{u,i}^T)^T$. Let $\boldsymbol{Z}^{obs} = (\boldsymbol{z}_1^{obs}, \cdots, \boldsymbol{z}_n^{obs})$ be the matrix of observations corresponding to the observed ordered categorical variables, where $\boldsymbol{z}_i^{obs} = (z_{i1}^{obs}, \cdots, z_{is}^{obs})^T$. Let $\boldsymbol{Y}^{rep} = (\boldsymbol{Y}_o^{rep}, \boldsymbol{Y}_u^{rep})$ be the replication of $\boldsymbol{Y}^{obs}$, and $\boldsymbol{Z}^{rep}$ be the replication of $\boldsymbol{Z}^{obs}$. Moreover, let $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$, $\boldsymbol{\Omega}_1 = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_n)$, and $\boldsymbol{\Omega}_2 = (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n)$; and let $\boldsymbol{\alpha}$ be the vector that contains all the unknown thresholds, and $\boldsymbol{\theta}$ be the vector that contains all the unknown parameters in $\boldsymbol{u}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\Lambda}_\omega, \boldsymbol{\Phi}$, and $\boldsymbol{\Psi}_\delta$. Besides, let Θ, Υ, Υ_y, and Ξ be the support spaces of $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$, $\boldsymbol{y}_{u,i}$, and $\boldsymbol{\xi}_i$, respectively, $i = 1, \cdots, n$.

3.3 L_v Measure for Nonlinear SEMs with Mixed Continuous and Ordered Categorical Data

L_v measure, originally proposed by Gelfand and Ghosh (1998), is based on exponential family. However, the distribution of ordered categorical data dose not belong to the exponential family. Therefore, we cannot use it directly. A method proposed by Chen et al. (2004) is used to transform the ordered categorical data into binary data in this chapter.

3.3.1 Definition of the L_v Measure

To handle the ordered categorical variable z_{ik} in M defined by equations (3.1) to (3.4), a new vector $\boldsymbol{z}_{ik}^*$ is introduced, where $\boldsymbol{z}_{ik}^* = (z_{ik,1}, \cdots, z_{ik,b_k+1})^T$, and the elements in $\boldsymbol{z}_{ik}^*$ are defined as follows: for $i = 1, \cdots, n$ and $k = 1, \cdots, s$,

$$z_{ik,j} = \begin{cases} 1, & \text{if } z_{ik} = j-1 \\ 0, & \text{otherwise} \end{cases}, \quad j = 1, \cdots, b_k + 1 \,. \tag{3.5}$$

From equation (3.4), we have

$$p(z_{ik,j} = 1) = p(z_{ik} = j-1) = p(\alpha_{k,j-1} < y_{u,ik} \leq \alpha_{k,j}) \triangleq p_{ik,j}, \tag{3.6}$$

and

$$p(z_{ik,j} = 0) = 1 - p(z_{ik,j} = 1) = 1 - p_{ik,j}. \tag{3.7}$$

It can be shown that $z_{ik,j}$ has a Bernoulli distribution with parameter $p_{ik,j}$, and the density function can be written as:

$$f(z_{ik,j}, p_{ik,j}) = p_{ik,j}^{z_{ik,j}}(1-p_{ik,j})^{1-z_{ik,j}} = \exp\{z_{ik,j}\theta_{ik,j} - b(\theta_{ik,j})\}, \tag{3.8}$$

where $\theta_{ik,j} = \log\frac{p_{ik,j}}{1-p_{ik,j}}$, and $b(\theta_{ik,j}) = \log(1+e^{\theta_{ik,j}})$. It belongs to the exponential family, and

$$E(z_{ik,j}) = \dot{b}(\theta_{ik,j}) = p_{ik,j}, \tag{3.9}$$

$$\mathrm{Var}(z_{ik,j}) = \ddot{b}(\theta_{ik,j}) = p_{ik,j}(1-p_{ik,j}), \tag{3.10}$$

where $\dot{b}(\cdot)$ and $\ddot{b}(\cdot)$ denote the first and second derivatives, respectively.

Now we consider the L_v measure for the ordered categorical variables defined in equation (3.4). According to equation (3.5), the elements z_{ik}^{obs} in $\boldsymbol{Z}^{obs}$ and z_{ik}^{rep} in $\boldsymbol{Z}^{rep}$ can be dichotomized as vectors $\boldsymbol{z}_{ik}^{obs*} = (z_{ik,1}^{obs}, \cdots, z_{ik,b_k+1}^{obs})$ and $\boldsymbol{z}_{ik}^{rep*} = (z_{ik,1}^{rep}, \cdots, z_{ik,b_k+1}^{rep})$, respectively. Let $\boldsymbol{Z}^{obs*} = \{z_{ik,j}^{obs*}; i = 1, \cdots, n;\ k = 1, \cdots, s; j = 1, \cdots, b_k+1\}$ and $\boldsymbol{Z}^{rep*} = \{z_{ik,j}^{rep*}; i = 1, \cdots, n;\ k = 1, \cdots, s; j = 1, \cdots, b_k+1\}$. Given observations $\boldsymbol{Y}_o^{obs}$ and $\boldsymbol{Z}^{obs}$, the quadratic loss L_v measure (see Gelfand and Ghosh, 1998; Ibrahim et al., 2001) is defined by

$$\begin{aligned} L_v^q(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) = \sum_{k=1}^{s}\sum_{i=1}^{n}[&tr\{\mathrm{Var}(\boldsymbol{z}_{ik}^{rep*}|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)\} \\ &+ v(\boldsymbol{z}_{ik}^{obs*} - \boldsymbol{\mu}_{ik}^*)^T(\boldsymbol{z}_{ik}^{obs*} - \boldsymbol{\mu}_{ik}^*)], \end{aligned} \tag{3.11}$$

where $\boldsymbol{\mu}_{ik}^* = E(\boldsymbol{z}_{ik}^{rep*}|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)$, in which the expectation is taken with respect to the posterior predictive distribution $p(\boldsymbol{z}_{ik}^{rep*}|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)$ defined by:

$$\begin{aligned} &p(\boldsymbol{z}_{ik}^{rep*} = \boldsymbol{e}_j|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) \\ &= \int_{\Omega_0} p(z_{ik}^{rep} = j-1, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, \boldsymbol{y}_{u,i}|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)d\boldsymbol{\theta}d\boldsymbol{\alpha}d\boldsymbol{\xi}_i d\boldsymbol{y}_{u,i}, \end{aligned} \tag{3.12}$$

where $\Omega_0 = \Theta \times \Upsilon \times \Xi \times \Upsilon_y$, $\boldsymbol{e}_j$ is a $(b_k+1)\times 1$ vector with 1 at the jth

element and 0 at the others, and

$$
\begin{aligned}
&p(z_{ik}^{rep} = j-1, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, \boldsymbol{y}_{u,i} | \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) \\
&= p(z_{ik}^{rep} = j-1 | \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, \boldsymbol{y}_{u,i}, M) p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, \boldsymbol{y}_{u,i} | \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) \qquad (3.13) \\
&= p(\alpha_{k,j-1} < y_{u,ik}^{rep} \leq \alpha_{k,j} | \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, M) p(\boldsymbol{\alpha}, \boldsymbol{\xi}_i, \boldsymbol{y}_{u,i} | \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M).
\end{aligned}
$$

It can be shown that equation (3.11) can be rewritten as:

$$
\begin{aligned}
&L_v^q(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) \\
&= \sum_{k=1}^{s} \sum_{i=1}^{n} \sum_{j=1}^{b_k+1} \left[\mathrm{Var}(z_{ik,j}^{rep*} | \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) + v(z_{ik,j}^{obs*} - \mu_{ik,j}^*)^2 \right], \qquad (3.14)
\end{aligned}
$$

where $\mathrm{Var}(z_{ik,j}^{rep*} | \boldsymbol{Z}^{obs}, M)$ is the jth diagonal element of the conditional covariance matrix $\mathrm{Var}(\boldsymbol{z}_{ik}^{rep*} | \boldsymbol{Z}^{obs}, M)$, and $\mu_{ik,j}^*$ is the jth element of the conditional expectation $\boldsymbol{\mu}_{ik}^*$. Therefore,

$$
\begin{aligned}
\mu_{ik,j}^* &= E(\boldsymbol{z}_{ik,j}^{rep*} | \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) \\
&= E\left[E(\boldsymbol{z}_{ik,j}^{rep*} | \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, \boldsymbol{y}_{u,i}, M) | \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M \right] \\
&= E\left[p(z_{ik,j}^{rep*} = 1 | \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, \boldsymbol{y}_{u,i}, M) | \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M \right] \qquad (3.15) \\
&= E\left[p(\alpha_{k,j-1} < y_{u,ik}^{rep} \leq \alpha_{k,j} | \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, M) | \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M \right].
\end{aligned}
$$

According to the definition of the proposed model, given $\boldsymbol{\theta}$ and $\boldsymbol{\xi}_i$, $\boldsymbol{y}_u$ has a normal distribution with mean $\boldsymbol{u}_u + \boldsymbol{\Lambda}_{u,\eta}(\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i)) + \boldsymbol{\Lambda}_{u,\xi}\boldsymbol{\xi}_i$, and covariance matrix $\boldsymbol{\Lambda}_{u,\eta}(\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Psi}_\delta(\boldsymbol{\Pi}_0^{-1})^T)\boldsymbol{\Lambda}_{u,\eta}^T + \boldsymbol{\Psi}_{\epsilon u}$, where $\boldsymbol{u}_u$, $\boldsymbol{\Lambda}_{u,\eta}$, $\boldsymbol{\Lambda}_{u,\xi}$, and $\boldsymbol{\Psi}_{\epsilon u}$ are the submatrices of $\boldsymbol{u}$, $\boldsymbol{\Lambda}_\eta$, $\boldsymbol{\Lambda}_\xi$, and $\boldsymbol{\Psi}_\epsilon$ corresponding to $\boldsymbol{y}_{u,i}$, respectively. Therefore,

$$
\begin{aligned}
&p(\alpha_{k,j-1} < y_{u,ik}^{rep} \leq \alpha_{k,j} | \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, M) \\
&= \Phi\left[\frac{\alpha_{k,j} - \left\{ u_{u,k} + \boldsymbol{\Lambda}_{u,\eta k}(\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i)) + \boldsymbol{\Lambda}_{u,\xi k}\boldsymbol{\xi}_i \right\}}{\boldsymbol{\Lambda}_{u,\eta k}(\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Psi}_\delta(\boldsymbol{\Pi}_0^{-1})^T)\boldsymbol{\Lambda}_{u,\eta k}^T + \psi_{\epsilon u k}} \right] \\
&- \Phi\left[\frac{\alpha_{k,j-1} - \left\{ u_{u,k} + \boldsymbol{\Lambda}_{u,\eta k}(\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i)) + \boldsymbol{\Lambda}_{u,\xi k}\boldsymbol{\xi}_i \right\}}{\boldsymbol{\Lambda}_{u,\eta k}(\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Psi}_\delta(\boldsymbol{\Pi}_0^{-1})^T)\boldsymbol{\Lambda}_{u,\eta k}^T + \psi_{\epsilon u k}} \right],
\end{aligned}
$$

where $u_{u,k}$ is the kth element of $\boldsymbol{u}_u$; $\boldsymbol{\Lambda}_{u,\eta k}$ and $\boldsymbol{\Lambda}_{u,\xi k}$ are the kth row of $\boldsymbol{\Lambda}_{u,\eta}$ and $\boldsymbol{\Lambda}_{u,\xi}$, respectively; $\phi_{\epsilon u k}$ is the kth diagonal element of $\boldsymbol{\Psi}_{\epsilon u}$; $\Phi(\cdot)$ is

the cumulative distribution function of the standard normal distribution. Thus, the conditional expectation required in equation (3.14) is obtained. Similarly, the conditional variance in equation (3.14) is given by

$$\begin{aligned}
&\mathrm{Var}(z_{ik,j}^{rep*}|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) \\
&= p(z_{ik,j}^{rep*} = 1|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)[1 - p(z_{ik,j}^{rep*} = 1|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)] \\
&= \mu_{ik,j}^*(1 - \mu_{ik,j}^*). \qquad (3.16)
\end{aligned}$$

It can be calculated easily after obtaining the conditional expectation $\mu_{ik,j}^*$.

For the observations $\boldsymbol{Y}_o^{obs}$ corresponding to the continuous variables, the L_v measure defined by equation (2.9) in Chapter 2 is applied. Therefore, given $(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$, the L_v measure for the model M is defined by

$$\begin{aligned}
&L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) = L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) + L_v^q(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) \\
&= \sum_{k=1}^{r}\sum_{i=1}^{n}\left[\mathrm{Var}(y_{ik}^{rep}|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) + v(\mu_{ik} - y_{ik}^{obs})^2\right] \\
&+ \sum_{k=1}^{s}\sum_{i=1}^{n}\sum_{j=1}^{b_k+1}\left[\mathrm{Var}(z_{ik,j}^{rep*}|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) + v(z_{ik,j}^{obs*} - \mu_{ik,j}^*)^2\right], \qquad (3.17)
\end{aligned}$$

where $\mu_{ik} = E(y_{ik}^{rep}|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$. Let $\boldsymbol{u}_o$, $\boldsymbol{\Lambda}_{o,\eta}, \boldsymbol{\Lambda}_{o,\xi}$ and $\boldsymbol{\Psi}_{\epsilon o}$ be the submatrices of $\boldsymbol{u}$, $\boldsymbol{\Lambda}_\eta, \boldsymbol{\Lambda}_\xi$, and $\boldsymbol{\Psi}_\epsilon$ corresponding to $\boldsymbol{y}_{o,i}^{obs}$, respectively. As discussed in Chapter 2,

$$\begin{aligned}
\mu_{ik} &= E\left[E(y_{ik}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M\right] \\
&= E\left[u_{o,k} + \boldsymbol{\Lambda}_{o,\eta k}\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i) + \boldsymbol{\Lambda}_{o,\xi k}\boldsymbol{\xi}_i|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M\right],
\end{aligned}$$

where $u_{o,k}$ is the kth element of $\boldsymbol{u}_o$; $\boldsymbol{\Lambda}_{o,\eta k}$ and $\boldsymbol{\Lambda}_{o,\xi k}$ are the kth row of $\boldsymbol{\Lambda}_{o,\eta}$ and $\boldsymbol{\Lambda}_{o,\xi}$, respectively. The conditional variance in the first summation of equation (3.17) is given by

$$\begin{aligned}
&\sigma_{ik} = \mathrm{Var}(y_{ik}^{rep}|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) \\
&= E\left[\mathrm{Var}(y_{ik}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, M)|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M\right] + \mathrm{Var}\left[E(y_{ik}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i)|\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M\right].
\end{aligned}$$

Due to the existence of intractable integrals in calculating the conditional expectation and variance, we can get neither a closed form of $L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs},$

$M)$, nor a closed form of $L_v^q(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)$. Therefore, MCMC methods are used to calculate the L_v measure defined by equations (3.14) and (3.17).

3.3.2 Computation of the L_v Measure

From its definition, L_v measure can be estimated with a sufficiently large number of random observations $\{\boldsymbol{\theta}^{(g)}, \boldsymbol{\alpha}^{(g)}, \boldsymbol{\Omega}^{(g)}, \boldsymbol{Y}_u^{(g)}; g = 1, \cdots, G\}$ generated from the posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{Y}_u | \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$. To generate this sample, Gibbs sampler (Gelman and Meng, 1998) given bellow is used. With current values $(\boldsymbol{\theta}^{(g)}, \boldsymbol{\alpha}^{(g)}, \boldsymbol{\Omega}^{(g)}, \boldsymbol{Y}_u^{(g)})$,

Step (a) Generate $\boldsymbol{\theta}^{(g+1)}$ from $p(\boldsymbol{\theta} | \boldsymbol{\alpha}^{(g)}, \boldsymbol{\Omega}^{(g)}, \boldsymbol{Y}_u^{(g)}, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$.

Step (b) Generate $\boldsymbol{\Omega}^{(g+1)}$ from $p(\boldsymbol{\Omega} | \boldsymbol{\theta}^{(g+1)}, \boldsymbol{\alpha}^{(g)}, \boldsymbol{Y}_u^{(g)}, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$.

Step (c) Generate $(\boldsymbol{\alpha}^{(g+1)}, \boldsymbol{Y}_u^{(g+1)})$ from $p(\boldsymbol{\alpha}, \boldsymbol{Y}_u | \boldsymbol{\theta}^{(g+1)}, \boldsymbol{\Omega}^{(g+1)}, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$.

After the convergence of the MCMC algorithm, the random observations are collected to get the estimation of the unknown parameters, the latent variables, and the L_v measure. The Bayesian estimation of unknown parameters and latent variables are given as follows:

$$\hat{\boldsymbol{\theta}} = \frac{1}{G}\sum_{g=1}^{G} \boldsymbol{\theta}^{(g)}, \quad \hat{\boldsymbol{\omega}}_i = \frac{1}{G}\sum_{g=1}^{G} \boldsymbol{\omega}_i^{(g)}, \quad \hat{\boldsymbol{\alpha}}_k = \frac{1}{G}\sum_{g=1}^{G} \boldsymbol{\alpha}_k^{(g)}$$

$$\widehat{\mathrm{Var}}(\boldsymbol{\theta}) = \frac{1}{G-1}\sum_{g=1}^{G} (\boldsymbol{\theta}^{(g)} - \hat{\boldsymbol{\theta}}^{(g)})(\boldsymbol{\theta}^{(g)} - \hat{\boldsymbol{\theta}}^{(g)})^T,$$

$$\widehat{\mathrm{Var}}(\boldsymbol{\omega}_i) = \frac{1}{G-1}\sum_{g=1}^{G} (\boldsymbol{\omega}_i^{(g)} - \hat{\boldsymbol{\omega}}_i{}^{(g)})(\boldsymbol{\omega}_i^{(g)} - \hat{\boldsymbol{\omega}}_i{}^{(g)})^T,$$

For $i = 1, \cdots, n$ and $k = 1, \cdots, r$, let

$$\mu_{ik}^{(g)} = u_{o,k}^{(g)} + \boldsymbol{\Lambda}_{o,\eta k}^{(g)} (\boldsymbol{\Pi}_0^{(g)})^{-1} \boldsymbol{\Gamma}^{(g)} \boldsymbol{F}(\boldsymbol{\xi}_i^{(g)}) + \boldsymbol{\Lambda}_{o,\xi k}^{(g)} \boldsymbol{\xi}_i^{(g)},$$

and

$$\widehat{\mu}_{ik} = \frac{1}{G}\sum_{g=1}^{G} \mu_{ik}^{(g)}, \quad \widehat{\sigma}_{ik} = \frac{1}{G}\sum_{g=1}^{G} \left[\psi_{\epsilon ok}^{(g+1)} + \left(\mu_{ik}^{(g+1)} \right)^2 \right] - (\widehat{\mu}_{ik})^2.$$

For $i = 1, \cdots, n$, $k = 1, \cdots, s$, and $j = 1, \cdots, b_k + 1$, let

$$\mu_{ik,j}^{*(g+1)} = \Phi[\frac{\alpha_{k,j}^{(g+1)} - \{u_{u,k}^{(g)} + \mathbf{\Lambda}_{u,\eta k}^{(g)}(\mathbf{\Pi}_0^{(g)})^{-1}\mathbf{\Gamma}^{(g)}\boldsymbol{F}(\boldsymbol{\xi}_i^{(g)}) + \mathbf{\Lambda}_{u,\xi k}^{(g)}\boldsymbol{\xi}_i^{(g)}\}}{\psi_{\epsilon uk}^{(g+1)}}]$$
$$- \Phi[\frac{\alpha_{k,j-1}^{(g+1)} - \{u_{u,k}^{(g)} + \mathbf{\Lambda}_{u,\eta k}^{(g)}(\mathbf{\Pi}_0^{(g)})^{-1}\mathbf{\Gamma}^{(g)}\boldsymbol{F}(\boldsymbol{\xi}_i^{(g)}) + \mathbf{\Lambda}_{u,\xi k}^{(g)}\boldsymbol{\xi}_i^{(g)}\}}{\psi_{\epsilon uk}^{(g+1)}}].$$

and

$$\widehat{\mu^*}_{ik,j} = \frac{1}{G}\sum_{g=1}^{G}\mu_{ik,j}^{*(g+1)}, \quad \widehat{\sigma^*}_{ik,j} = \widehat{\mu}_{ik,j}^*\left(1 - \widehat{\mu}_{ik,j}^*\right).$$

Then the estimation of the L_v measure is given by

$$\begin{aligned} L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) = &\sum_{k=1}^{r}\sum_{i=1}^{n}\left[\widehat{\sigma}_{ik} + v(\widehat{\mu}_{ik} - y_{ik}^{obs})^2\right] \\ &+ \sum_{k=1}^{s}\sum_{i=1}^{n}\sum_{j=1}^{b_k+1}\left[\widehat{\sigma^*}_{ik,j} + v(z_{ik,j}^{obs*} - \widehat{\mu}_{ik,j}^*)^2\right], \end{aligned} \tag{3.18}$$

Now we consider the conditional distributions required in steps (a), (b), and (c) in implementing the Gibbs sampler.

3.3.3 Full Conditional Distributions

To obtain the conditional distribution $p(\boldsymbol{\theta}|\ \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, \mathbf{\Omega},\ \boldsymbol{Y}_u)$ required in **Step (a)**, the prior distributions of the parameters in $\boldsymbol{\theta}$ are needed. According to the suggestion given by Lee and Zhu (2000), the following commonly used conjugate prior distributions are specified:

$$\begin{aligned} &p(\boldsymbol{u}) \triangleq N[\boldsymbol{u}_0, \mathbf{\Sigma}_0], \quad p(\mathbf{\Phi}) \triangleq IW_{q_2}[\boldsymbol{R}_0, \rho_0], \\ &p(\psi_{\delta k}^{-1}) \triangleq \text{Gamma}[\alpha_{0\delta k}, \beta_{0\delta k}], \quad p(\psi_{\epsilon k}^{-1}) \triangleq \text{Gamma}[\alpha_{0\epsilon k}, \beta_{0\epsilon k}], \\ &p(\mathbf{\Lambda}_k|\psi_{\epsilon k}) \triangleq N[\mathbf{\Lambda}_{0k}, \psi_{\epsilon k}\boldsymbol{H}_{0k}], \quad p(\mathbf{\Lambda}_{\omega k}|\psi_{\delta k}) \triangleq N[\mathbf{\Lambda}_{0\omega k}, \psi_{\delta k}\boldsymbol{H}_{0\omega k}], \end{aligned} \tag{3.19}$$

where $\psi_{\epsilon k}$ and $\psi_{\delta k}$ are the kth diagonal elements of $\mathbf{\Psi}_\epsilon$ and $\mathbf{\Psi}_\delta$, respectively. $\mathbf{\Lambda}_k$ and $\mathbf{\Lambda}_{\omega k}$ are the kth row of $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\omega$, respectively; $\boldsymbol{u}_0$, $\alpha_{0\epsilon k}$, $\beta_{0\epsilon k}$, $\mathbf{\Lambda}_{0k}$, $\alpha_{0\delta k}$, $\beta_{0\delta k}$, $\mathbf{\Lambda}_{0\omega k}$, and ρ_0, and positive definite matrices $\mathbf{\Sigma}_0, \boldsymbol{H}_{0k}, \boldsymbol{H}_{0\omega k}$, and $\boldsymbol{R}_0$, are hyperparameters whose values are given by the prior information. It

can be shown that conditional distributions of the components in $\boldsymbol{\theta}$ are the familiar normal, Gamma, and inverted Wishart distributions. Generating observations from these distributions are simple and fast.

For the conditional distribution $p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \boldsymbol{Y}_u, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$ required in **Step (b)**,

$$\begin{aligned} p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \boldsymbol{Y}_u, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}) &= \prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\boldsymbol{\theta}, \boldsymbol{y}_i) \\ &\propto \prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\omega}_i, \boldsymbol{\theta}) p(\boldsymbol{\eta}_i|\boldsymbol{\xi}_i, \boldsymbol{\theta}) p(\boldsymbol{\xi}_i|\boldsymbol{\theta}). \quad (3.20) \end{aligned}$$

According to the definition of the model,

$$\begin{aligned} p(\boldsymbol{\omega}_i|\boldsymbol{\theta}) \propto \exp\Bigg\{ & -\frac{1}{2}\boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i - \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_\epsilon^{-1} (\boldsymbol{y}_i - \boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) \\ & -\frac{1}{2}\left[\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{\omega}_i)\right]^T \boldsymbol{\Psi}_\delta^{-1} \left[\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \boldsymbol{G}(\boldsymbol{\omega}_i)\right] \Bigg\}. \quad (3.21) \end{aligned}$$

Thus, the required conditional distribution in **Step (b)** can be obtained through equations (3.20) and (3.21).

Finally, consider the joint conditional distribution of $(\boldsymbol{\alpha}, \boldsymbol{Y}_u)$, given $\boldsymbol{Y}_o^{obs}$, $\boldsymbol{Z}^{obs}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\theta}$. The following non-informative prior distribution of $\boldsymbol{\alpha}$ is used: for $k = 1, \cdots, s$,

$$p(\boldsymbol{\alpha}_k) = p(\alpha_{k,2}, \cdots, \alpha_{k,b_k-1}) \propto C, \qquad \text{for } \alpha_{k,2} < \cdots < \alpha_{k,b_k-1},$$

where C is a constant. As $\boldsymbol{\Psi}_\epsilon$ is diagonal, $(\boldsymbol{\alpha}, \boldsymbol{Y}_u)$ is conditionally independent of $\boldsymbol{Y}_o^{obs}$ and $\boldsymbol{Z}^{obs}$ given $\boldsymbol{\Omega}$ and $\boldsymbol{\theta}$, and

$$\begin{aligned} & p(\boldsymbol{\alpha}_k, \boldsymbol{y}_{u(k)}|\boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}) \\ &= p(\boldsymbol{\alpha}_k|\boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{y}_{u(k)}) p(\boldsymbol{y}_{u(k)}|\boldsymbol{\theta}, \boldsymbol{\alpha}_k, \boldsymbol{\Omega}, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}) \\ &= p(\boldsymbol{\alpha}_k|\boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{y}_{u(k)}) \prod_{i=1}^{n} p(y_{u,ik}|\boldsymbol{\theta}, \boldsymbol{\alpha}_k, \boldsymbol{\omega}_i, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}), \end{aligned}$$

where $\boldsymbol{y}_{u(k)} = (y_{u,1k}, \cdots, y_{u,nk})$, and

$$p(\boldsymbol{\alpha}_k|\boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{y}_{u(k)}) \propto \prod_{i=1}^{n} \left[\Phi\left(\frac{\alpha_{k,z_{ik}+1} - u_{u,k} - \boldsymbol{\Lambda}_{u,k}^T \boldsymbol{\omega}_i}{\psi_{\epsilon uk}^{1/2}} \right) - \Phi\left(\frac{\alpha_{k,z_{ik}} - u_{u,k} - \boldsymbol{\Lambda}_{u,k}^T \boldsymbol{\omega}_i}{\psi_{\epsilon uk}^{1/2}} \right) \right]. \tag{3.22}$$

and

$$p(y_{u,ik}|\boldsymbol{\alpha}_k, y_{o,ik}^{obs}, \boldsymbol{\Omega}, \boldsymbol{\theta}) \triangleq N(u_{u,k} + \boldsymbol{\Lambda}_{u,k}^T \boldsymbol{\omega}_i, \psi_{\epsilon uk}) I_{(\alpha_{k,z_{k,i}}, \alpha_{k,z_{ik}+1})}(y_{u,ik}). \tag{3.23}$$

From equations (3.22) and (3.23), the conditional distribution required in **Step (c)** is obtained.

The conditional distributions involved in $p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \boldsymbol{Y}_u, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$ and $p(\boldsymbol{\alpha}, \boldsymbol{Y}_u|\boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$ are non-standard and complex. The Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) is used to simulate observations from them. Thus, the proposed MCMC algorithm for posterior simulation is a hybrid algorithm that combines the Gibbs sampler and the MH algorithm.

3.3.4 Calibration Distribution

Let $L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)$ denote the L_v measure for the candidate model M_c, and $L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_t)$ denote the L_v measure for the true model M_t. Then the difference between the L_v measures of M_c and M_t is defined as

$$D_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c) \equiv L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c) - L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_t), \tag{3.24}$$

and the calibration distribution for model M_c, denoted by PL_c, is defined as the marginal distribution of $D_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)$, computed with respect to the prior predictive distribution of $(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$ under the true model M_t:

$$p_t(\boldsymbol{Z}^{obs}, \boldsymbol{Y}_o^{obs}) = \int_{\Theta \times \Upsilon} p(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}|\boldsymbol{\theta}, \boldsymbol{\alpha}, M_t) p(\boldsymbol{\theta}|M_t) p(\boldsymbol{\alpha}|M_t) d\boldsymbol{\theta} d\boldsymbol{\alpha}, \tag{3.25}$$

where

$$p(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}|\boldsymbol{\theta}, \boldsymbol{\alpha}, M_t)$$
$$= \int_{\Xi \times \Upsilon_y} p(\boldsymbol{Z}^{obs}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}_2, \boldsymbol{Y}_u, M_t) p(\boldsymbol{Y}_o^{obs}|\boldsymbol{\theta}, \boldsymbol{\Omega}_2, M_t) d\boldsymbol{\Omega}_2 d\boldsymbol{Y}_u.$$

Thus the calibration distribution PL_c is defined by

$$PL_c \equiv p(D_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)). \tag{3.26}$$

From this definition, it is impossible to obtain a closed form of the calibration distribution. As discussed in Chapter 2, MCMC methods are used to estimate the calibration distribution. As the true model M_t is required in calculating the calibration distribution, we will treat the model with the smallest value of the L_v measure as the true model in real applications. The procedure is given as follows:

(a) Generate $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\Omega}}_2, \tilde{\boldsymbol{Y}}_u)$ from $p(\boldsymbol{\Omega}_2, \boldsymbol{Y}_u|\boldsymbol{\theta}, \tilde{\boldsymbol{\alpha}}, M_t) p(\boldsymbol{\theta}|M_t) p(\boldsymbol{\alpha}|M_t)$, where $p(\boldsymbol{\theta}|M_t)$ and $p(\boldsymbol{\alpha}|M_t)$ are proper prior distributions of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ under M_t,

(b) Generate a pseudo observation $(\tilde{\boldsymbol{Y}}_o, \tilde{\boldsymbol{Z}})$ from $p(\boldsymbol{Y}, \boldsymbol{Z}|\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\Omega}}_2, \tilde{\boldsymbol{Y}}_u)$,

(c) Set $(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}) = (\tilde{\boldsymbol{Y}}_o, \tilde{\boldsymbol{Z}})$, and use the method described above to obtain the estimates of $L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)$ and $L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_t)$, then $D_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)$.

Repeat (a), (b), and (c) G times, we can obtain a sample $\{D_v^{(g)}(\boldsymbol{Z}^{obs}, \boldsymbol{Y}_o^{obs}, M_c), g = 1, \cdots, G\}$. Based on this sample, we can estimate the calibration distribution PL_c via the kernel density estimation method (Silverman, 1986; Sheather and Jones, 1991; Scott, 1992). The summaries of PL_c can be obtained easily through the methods given in Chapter 2.

3.4 A Simulation Study

Model Setting

In this section, a simulation study is presented to evaluate the performance of the L_v measure. The observations $\{\boldsymbol{y}_{o,i}, \boldsymbol{z}_i; i = 1, \cdots, n\}$ are generated from the following structural equation model:

$$M_0 : \ \boldsymbol{y}_i = \boldsymbol{u} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \text{ and } \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i1}\xi_{i2} + \delta_i,$$

where $\boldsymbol{y}_i = (\boldsymbol{y}_{o,i}^T, \boldsymbol{y}_{u,i}^T)^T$ is a 9×1 vector, and $\boldsymbol{z}_i(4 \times 1)$ which corresponds to $\boldsymbol{y}_{u,i}(4 \times 1)$ is generated from equation (3.4) through threshold $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_4)$. $\boldsymbol{\omega}_i = (\eta_i, \xi_{i1}, \xi_{i2})$ is a 3×1 vector of latent variables. The specification of $\boldsymbol{\Lambda}$ is

$$\boldsymbol{\Lambda}^T = \begin{pmatrix} 1.0^* & \lambda_{21} & \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{52} & \lambda_{62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{83} & \lambda_{93} \end{pmatrix},$$

where the elements with asterisks are fixed. True values for the unknown parameter are given as: $\lambda_{21} = \lambda_{31} = \lambda_{83} = \lambda_{93} = 0.8$, $\lambda_{52} = \lambda_{62} = 0.7$, $u_1 = \cdots = u_9 = 0.0$, $\gamma_1 = \gamma_2 = 0.6, \gamma_3 = -0.5$; $\phi_{11} = \phi_{22} = 1.0$, $\phi_{12} = \phi_{21} = 0.2$, $\psi_{\epsilon 1} = \cdots = \psi_{\epsilon 9} = 0.5$, $\psi_\delta = 0.5$, and $\boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_4 = (-1.0^*, -0.6, 0.6, 1.0^*)^T$, in which -1.0 and 1.0 are fixed. Four competing models with the same measurement equation as M_0 but different structural equations are considered. The structural equations are given as follows:

$$\begin{aligned} M_1 &: \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i1}^2 + \gamma_4\xi_{i2}^2 + \gamma_5\xi_{i1}\xi_{i2} + \delta_i, \\ M_2 &: \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \delta_i, \\ M_3 &: \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i1}^2 + \delta_i, \\ M_4 &: \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i2}^2 + \delta_i. \end{aligned}$$

To compare different model selection methods, results related to L_v measure, Bayes factor, and DIC will be presented. Similar as in Chapter 2, to

study the effect of sample size and prior inputs of the parameters, two different types of prior inputs, Prior I and Prior II, and three different sample sizes, n=150, 300 and 600, are considered.

Prior I: In the prior distributions of $\boldsymbol{u}$, $\boldsymbol{\Lambda}_\epsilon$, and $\boldsymbol{\Lambda}_\omega$, the means of the normal distributions are taken as the true values, and the covariance matrices are taken to be the identity matrices with appropriate dimensions. In the prior distribution of $\boldsymbol{\Phi}$, ρ_0, and $\boldsymbol{R}_0$ in the Wishart distribution are taken to be 7 and 4 times of $\boldsymbol{\Phi}_0$, respectively, where $\boldsymbol{\Phi}_0$ is the matrix with true values of ϕ_{11}, ϕ_{21} and ϕ_{22}. In the prior distributions of ψ_δ and $\psi_{\epsilon k}$, the hyperparameters in the Gamma distributions are taken to be $\alpha_{0k} = \alpha_{0\delta} = 9$ and $\beta_{0k} = \beta_{0\delta} = 4$.

Prior II: The means of the normal distributions are all zero, the covariance matrices are equal to four times of the identity matrices with appropriate dimensions, $\rho_0 = 2$, $\boldsymbol{R}_0$ is the identity matrix, $\alpha_{0k} = \alpha_{0\delta} = 4$ and $\beta_{0k} = \beta_{0\delta} = 5$.

L_v measure for model selection

In this simulation study, 100 replications are conducted to get the estimation of unknown parameters, the L_v measure, and the calibration distribution. Results of the calibration summaries are given in Table 3.1, where mean($L_{0.5}$) denotes the mean value of the L_v measure with $v = 0.5$, $\mu_{0.5}$, $SD_{0.5}$, and 95% HPD denote the mean, the standard deviation, and the 95% HPD interval of the calibration distribution of the corresponding model with $v = 0.5$, respectively. From this table, under each case, $\mu_{0.5}$ for M_1 is close to zero, and the corresponding 95% HPD interval contains zero, we conclude that the performances of models M_0 and M_1 are similar. According to the parsimonious principle, M_0 is selected. Compared with $\mu_{0.5}$ for M_1, $\mu_{0.5}$ for models M_2, M_3, and M_4 are much larger than zero. We conclude that the performance of M_0 is better than M_2, M_3, and M_4 on average. Apart from the means, we also see that when the sample size is

small, some of the 95% HPD intervals for M_2 to M_4 contain zero. However, considering the large values of the corresponding upper bound, the lower bounds are relatively close to zero, which means that M_0 is favorable. When sample size $n = 600$, it is obvious that M_0 performs much better than M_2, M_3, and M_4. Thus, M_0 is selected under each case.

The calibration distributions under Prior I and Prior II are given in Figures 3.1(a) to 3.1(c) and Figures 3.2(a) to 3.2(c), respectively. The same conclusion as above is obtained. Therefore, the true model can be consistently selected under the considered different types of prior inputs and different sample sizes. Besides, when sample size increases, the centers of the calibration distributions for M_2, M_3, and M_4 become further apart from zero, but that for M_1 changes little.

Bayes Factor for model selection

To calculate the Bayes factor, path sampling (Gelman and Meng, 1998) is used. M_0 and the four candidate models $M_1, \cdots, M_4$ can be linked up similarly as Chapter 2 with the parameter t, where $t \in [0, 1]$. In the path sampling procedure, we take $S = 20$ grids in $[0, 1]$. Based on the previous analysis, for each $t_{(s)}$, we take a burn-in phase of 4000 iterations, and further collect 2000 observations in computing the logarithm Bayes factor. The estimated logarithm Bayes factors are presented in Table 3.2. From this table, all the logarithm Bayes factors are negative, which consistently select M_0. Furthermore, the estimated logarithm Bayes factors have similar values under Prior I and Prior II. Therefore, the same conclusion can be drawn under these two different prior inputs.

Model selection using DIC and WinBUGS

In this part, model selection based on DIC is discussed. DIC value can be obtained similarly as before through WinBUGS. We use a burn-in phase of 4000 iterations, and further collect 2000 samples to calculate DIC. The

DIC values under Prior I with different sample sizes are given in Table 3.3. From this table, M_0 with the smallest DIC value is selected for each given sample size.

3.5 A Real Example

Measures of quality of life (QOL) and/or health-related QOL have great value for clinical work, and the planning and evaluation of health care. A Bayesian method for analyzing a common QOL data with ordered categorical items has been discussed in Lee (2007). The aim of this section is to apply L_v measure to model selection in the analysis of this QOL data. The instrument WHOQOL-100 (see Power et al., 1999) in measuring QOL was established to evaluate four latent constructs: physical health, psychological health, social relationships, and environment. In the instrument, Q3 to Q9 measure 'physical health', Q10 to Q15 measure 'psychological health', Q16 to Q18 measure 'social relationships', and the last eight items (Q19 to Q26) measure 'environment'. In addition to the 24 ordered categorical items, the instrument also includes two ordered categorical items, the overall QOL (Q1) and the health-related QOL (Q2), giving a total of 26 items. All of the items are measured with a 5-point scale (1 = 'not at all/very dissatisfied'; 2 = 'a little/dissatisfied'; 3 = 'moderate/neither'; 4 ='very much/satisfied'; 5 = 'extremely/very satisfied'). The sample size of the whole data set is extremely large. To illustrate the performance of L_v measure, we only analyze a synthetic data set with sample size n = 338. Similar as in Lee (2007), we compare a SEM (M_1) with four exogenous latent variables with another SEM (M_2) with three exogenous latent variables. The measurement equation of M_1 is defined by

$$\boldsymbol{y} = \boldsymbol{\Lambda}_1 \boldsymbol{\omega}_1 + \boldsymbol{\epsilon}, \tag{3.27}$$

where $\boldsymbol{\omega}_1 = (\eta, \xi_1, \xi_2, \xi_3, \xi_4)^T$, $\boldsymbol{\epsilon} \sim N[0, \boldsymbol{\Psi}_{\epsilon 1}]$, and

$$\boldsymbol{\Lambda}_1^T = \begin{bmatrix} 1 & \lambda_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \lambda_{42} & \cdots & \lambda_{92} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{11,3} & \cdots & \lambda_{15,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{17,4} & \lambda_{18,4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{20,5} & \cdots & \lambda_{26,5} \end{bmatrix}.$$

The structural equation of M_1 is given by

$$\eta = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_3 + \gamma_4 \xi_4 + \delta, \tag{3.28}$$

where $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_4)^T$ and δ are independently distributed as $N[\mathbf{0}, \boldsymbol{\Phi}_1]$ and $N[0, \sigma_{1\delta}^2]$, respectively. The measurement equation of M_2 is defined by

$$\boldsymbol{y} = \boldsymbol{\Lambda}_2 \boldsymbol{\omega}_2 + \boldsymbol{\epsilon}, \tag{3.29}$$

where $\boldsymbol{\omega}_2 = (\eta, \xi_1, \xi_2, \xi_3)^T$, $\boldsymbol{\epsilon} \sim N[0, \boldsymbol{\Psi}_{\epsilon 2}]$, and

$$\boldsymbol{\Lambda}_2^T = \begin{bmatrix} 1 & \lambda_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \lambda_{42} & \cdots & \lambda_{92} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{11,3} & \cdots & \lambda_{15,3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{17,4} & \cdots & \lambda_{26,4} \end{bmatrix}.$$

The structural equation of M_2 is given by

$$\eta = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_3 + \delta, \tag{3.30}$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^T$ and δ are independently distributed as $N[\mathbf{0}, \boldsymbol{\Phi}_2]$ and $N[0, \sigma_{2\delta}^2]$, respectively.

In the above two models, $\boldsymbol{y}$ is the underlying vector of manifest variables, which corresponds to the observation $\boldsymbol{z}$. The relationship between $\boldsymbol{z}$ and $\boldsymbol{y}$ is defined by equation (3.4). The threshold are given by $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{26})^T$, where $\boldsymbol{\alpha}_k = (\alpha_{k1}, \alpha_{k2}, \cdots, \alpha_{k5}, \alpha_{k6})$, $\alpha_{k1} = -\infty, \alpha_{k6} = \infty$. For identification, some elements of the thresholds will be fixed at certain values. Here the standard normal distribution $N[0, 1]$ is applied to y_k, and then α_{k2} and α_{k5} can be fixed according to the cumulative frequencies of the ordered categorical items, see Lee (2007) for more details.

L_v measure for model selection

To calculate the L_v measure, the conjugate prior distributions are used. The hyperparameter values corresponding to the prior distributions of the unknown loadings in $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are all taken to be 0.8; those corresponding to $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ are $(0.6, 0.6, 0.4, 0.4)$; those corresponding to $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are $\rho_1 = \rho_2 = 30$, $\boldsymbol{R}_{01}^{-1} = 8\boldsymbol{I}_4$, and $\boldsymbol{R}_{02}^{-1} = 8\boldsymbol{I}_3$, respectively; $\boldsymbol{H}_{0k1} = \boldsymbol{H}_{0k2} = 0.25\boldsymbol{I}_{26}$, $\boldsymbol{H}_{0\omega k1} = 0.25\boldsymbol{I}_4$, and $\boldsymbol{H}_{0\omega k2} = 0.25\boldsymbol{I}_3$, where $\boldsymbol{I}_d$ denotes an identity matrix with dimension d; $\alpha_{0\epsilon k1} = \alpha_{0\delta k1} = \alpha_{0\epsilon k2} = \alpha_{0\delta k2} = 10$ and $\beta_{0\epsilon k1} = \beta_{0\delta k1} = \beta_{0\epsilon k2} = \beta_{0\delta k2} = 8$. In the Gibbs sampling in computing the L_v measure and the estimation of unknown parameters, we take $J = 2000$ observations after a burn-in phase of 4000 iterations. L_v measure is 7273.01 for M_1 and 7343.826 for M_2. As the value of the L_v measure of M_1 is less than that of M_2, M_1 is selected. To obtain the calibration distribution, 100 data sets are generated based on M_1 under Prior I. The calibration distribution summaries are given in Table 3.4, and the density of calibration distribution is given in Figure 3.3. We see that the mean of the difference between the L_v measures of M_1 and M_2 is larger than zero, and the 95% HPD interval dose not include zero. Therefore, M_1 will be selected. The estimation of the the unknown parameters are given in Table 3.5.

Bayes factor for model selection

To compare M_1 and M_2 by using Bayes factor, path sampling (Gelman and Meng, 1998) is applied. First, we will compare model M_1 with the following model M_0:

$$M_0 : \ \boldsymbol{y} = \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} \sim N[\mathbf{0}, \mathbf{\Psi}_\epsilon]$, and $\mathbf{\Psi}_\epsilon$ is a diagonal matrix. We obtain $\log B_{10} = 81.36$. Similarly, M_2 and M_0 can be compared via the path sampling procedure, and $\log B_{20} = 57.85$, which means that M_1 and M_2 are both better than M_0. Furthermore, from the above result, $\log B_{12}$ is equal to 23.51. Therefore, M_1 is selected.

WinBUGS and DIC for model selection

For a SEM with ordered categorical variables, the software WinBUGs can produce the Bayesian estimates of the structural parameters and latent variables in the model, as well as the DIC value for model selection. In this example, DIC value is 19532.8 for model M_1, and 19609.3 for model M_2. Therefore, M_1 is selected.

3.6 Discussion

From the numerical studies in this chapter, the same conclusion in model selection is obtained by using the L_v measure, Bayes factor, and DIC. As mentioned in Chapter 2, the computation of the L_v measure is much more efficient than Bayes factor. Besides, the L_v measure is more reasonable than DIC because it incorporates the calibration distribution to further helping making decision in model selection. In this chapter, we propose L_v measure for nonlinear SEMs with mixed continuous and ordered categorical data. The simulation study demonstrates that our proposed method performs well. A quality of life data is analyzed for illustration.

Table 3.1: Mean values of the L_v measure and calibration summaries for simulation study

	Sample size	Model	mean($L_{0.5}$)	$\mu_{0.5}$	$SD_{0.5}$	95% HPD
Prior I	n=150	M_0	1172.887	-	-	-
		M_1	1176.119	3.231	13.371	(-13.263, 44.183)
		M_2	1195.186	22.298	13.444	(-0.660, 49.418)
		M_3	1193.872	20.985	14.777	(-6.415, 47.962)
		M_4	1192.788	19.901	11.664	(3.380, 45.548)
	n=300	M_0	2327.844	-	-	-
		M_1	2328.963	1.118	10.463	(-14.350, 22.744)
		M_2	2372.570	44.726	21.968	(5.489, 79.529)
		M_3	2369.000	41.155	25.904	(8.724, 84.179)
		M_4	2374.842	46.998	34.075	(2.470, 88.519)
	n=600	M_0	4645.018	-	-	-
		M_1	4644.271	-0.746	14.661	(-22.804, 36.000)
		M_2	4739.627	94.609	37.701	(28.232, 178.181)
		M_3	4732.777	87.760	34.010	(31.140, 162.030)
		M_4	4732.938	87.920	31.939	(30.122, 150.164)
Prior II	n=150	M_0	1252.040	-	-	-
		M_1	1253.551	1.511	12.758	(-17.913, 36.100)
		M_2	1271.365	19.325	13.001	(0.908, 52.612)
		M_3	1271.588	19.548	14.288	(-7.796, 50.790)
		M_4	1271.512	19.471	13.826	(-3.864, 47.642)
	n=300	M_0	2411.128	-	-	-
		M_1	2413.038	1.910	20.953	(-19.854, 77.502)
		M_2	2455.364	44.237	22.139	(-4.492, 89.980)
		M_3	2450.996	39.869	21.143	(-0.313, 83.849)
		M_4	2452.959	41.832	22.702	(5.707, 100.516)
	n=600	M_0	4730.851	-	-	-
		M_1	4735.537	4.685	24.557	(-20.688, 81.630)
		M_2	4818.795	87.943	32.794	(34.767, 158.430)
		M_3	4812.197	81.345	28.582	(24.479, 130.711)
		M_4	4818.343	87.492	35.130	(23.915, 158.075)

Table 3.2: Logarithm Bayes factor for simulation study

	Logarithm Bayes factor	n=150	n=300	n=600
Prior I	$\log B_{10}$	-1.89	-2.56	-3.25
	$\log B_{20}$	-6.05	-13.43	-29.65
	$\log B_{30}$	-6.78	-13.548	-29.386
	$\log B_{40}$	-6.77	-13.886	-30.02
Prior II	$\log B_{10}$	-2.05	-2.63	-3.29
	$\log B_{20}$	-5.43	-12.56	-28.54
	$\log B_{30}$	-6.11	-12.63	-28.36
	$\log B_{40}$	-6.18	-13.02	-27.85

Table 3.3: DIC values for simulation study

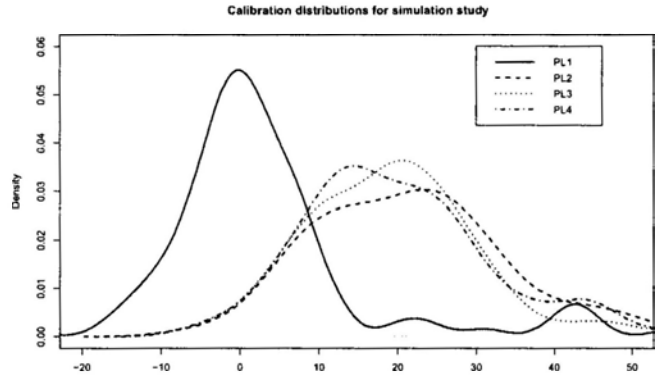
model	n=150	n=300	n=600
M_0	3550.29	7097.44	14200.96
M_1	3556.84	7118.46	14265.34
M_2	3606.42	7169.57	14734.12
M_3	3614.07	7192.41	14362.13
M_4	3600.47	7185.42	14375.42

Table 3.4: Calibration summaries for real example ($\nu = 0.5$)

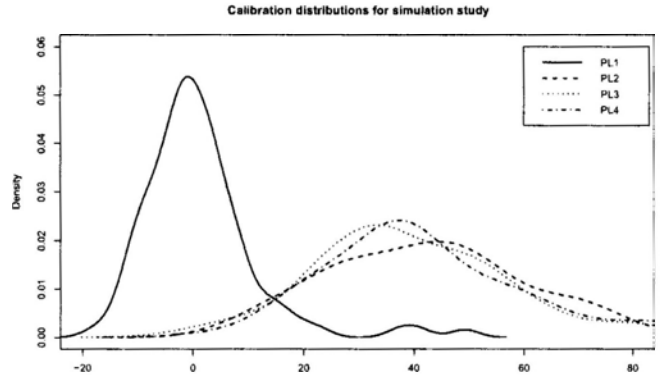
Model	$\mu_{0.5}(D)$	$SD(D_{0.5})$	95% HPD
M_2	95.6	17.896	(68.783, 135.124)

Table 3.5: Bayesian estimates of unknown parameters in M_1 for real example

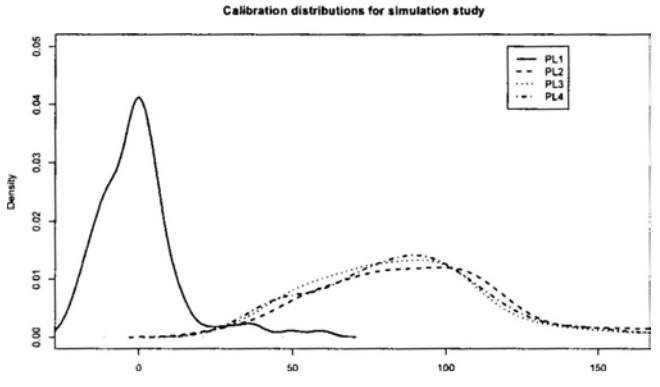
Parameter	EST	SD	Parameter	EST	SD	Parameter	EST	SD
λ_{21}	0.85	0.07	ψ_{δ}	0.25	0.03	$\psi_{\epsilon 17}$	0.96	0.09
λ_{42}	0.91	0.09	γ_1	0.76	0.09	$\psi_{\epsilon 18}$	0.52	0.06
λ_{52}	1.06	0.08	γ_2	0.37	0.1	$\psi_{\epsilon 19}$	0.53	0.06
λ_{62}	1.14	0.09	γ_3	0.14	0.11	$\psi_{\epsilon 20}$	0.67	0.07
λ_{72}	0.79	0.09	γ_4	-0.03	0.11	$\psi_{\epsilon 21}$	0.7	0.07
λ_{82}	1.26	0.08	$\psi_{\epsilon 1}$	0.39	0.05	$\psi_{\epsilon 22}$	0.7	0.07
λ_{92}	1.14	0.08	$\psi_{\epsilon 2}$	0.42	0.05	$\psi_{\epsilon 23}$	0.74	0.07
$\lambda_{11,3}$	0.8	0.09	$\psi_{\epsilon 3}$	0.62	0.07	$\psi_{\epsilon 24}$	0.57	0.06
$\lambda_{12,3}$	0.72	0.08	$\psi_{\epsilon 4}$	0.61	0.07	$\psi_{\epsilon 25}$	0.71	0.07
$\lambda_{13,3}$	0.75	0.09	$\psi_{\epsilon 5}$	0.46	0.05	$\psi_{\epsilon 26}$	0.66	0.07
$\lambda_{14,3}$	1	0.08	$\psi_{\epsilon 6}$	0.4	0.05	ϕ_{11}	0.49	0.06
$\lambda_{15,3}$	0.86	0.08	$\psi_{\epsilon 7}$	0.7	0.06	ϕ_{12}	0.35	0.04
$\lambda_{17,4}$	0.28	0.09	$\psi_{\epsilon 8}$	0.28	0.03	ϕ_{13}	0.22	0.04
$\lambda_{18,4}$	0.95	0.1	$\psi_{\epsilon 9}$	0.39	0.04	ϕ_{14}	0.31	0.04
$\lambda_{20,5}$	0.8	0.08	$\psi_{\epsilon 10}$	0.47	0.05	ϕ_{22}	0.58	0.07
$\lambda_{21,5}$	0.77	0.09	$\psi_{\epsilon 11}$	0.65	0.07	ϕ_{23}	0.38	0.05
$\lambda_{22,5}$	0.76	0.09	$\psi_{\epsilon 12}$	0.71	0.07	ϕ_{24}	0.39	0.05
$\lambda_{23,5}$	0.71	0.09	$\psi_{\epsilon 13}$	0.7	0.07	ϕ_{33}	0.59	0.08
$\lambda_{24,5}$	0.97	0.1	$\psi_{\epsilon 14}$	0.45	0.05	ϕ_{34}	0.38	0.05
$\lambda_{25,5}$	0.77	0.09	$\psi_{\epsilon 15}$	0.57	0.06	ϕ_{44}	0.54	0.07
$\lambda_{26,5}$	0.84	0.1	$\psi_{\epsilon 16}$	0.46	0.06			



(a) n=150

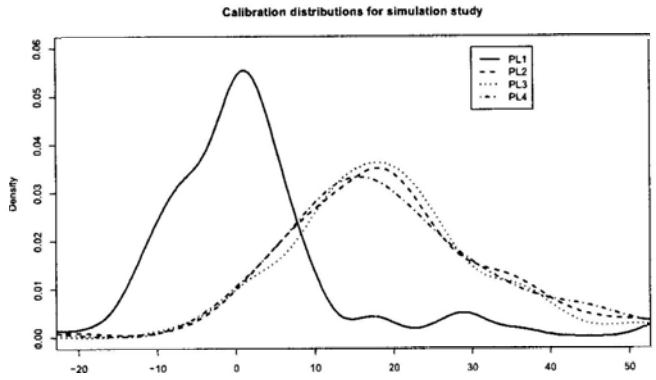


(b) n=300

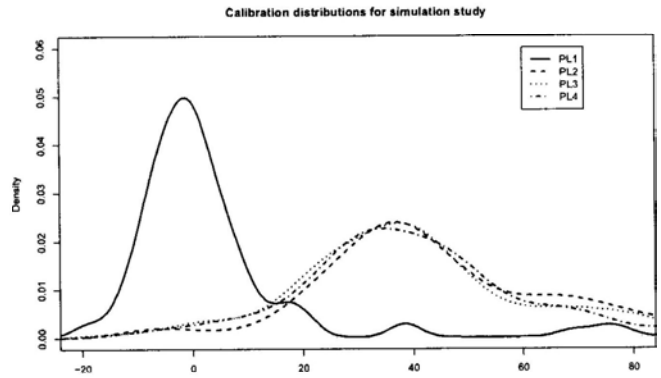


(c) n=600

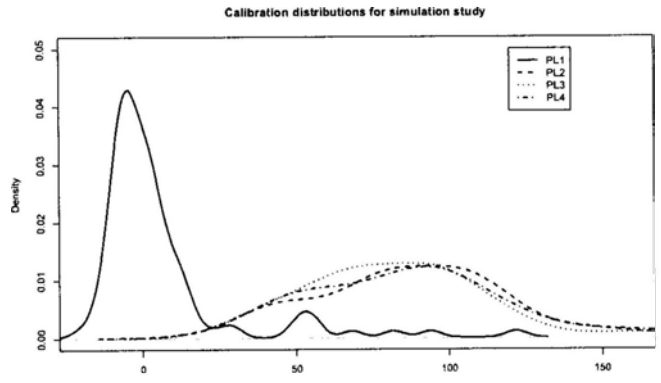
Figure 3.1: Calibration distributions under Prior I



(a) n=150



(b) n=300



(c) n=600

Figure 3.2: Calibration distributions under Prior II

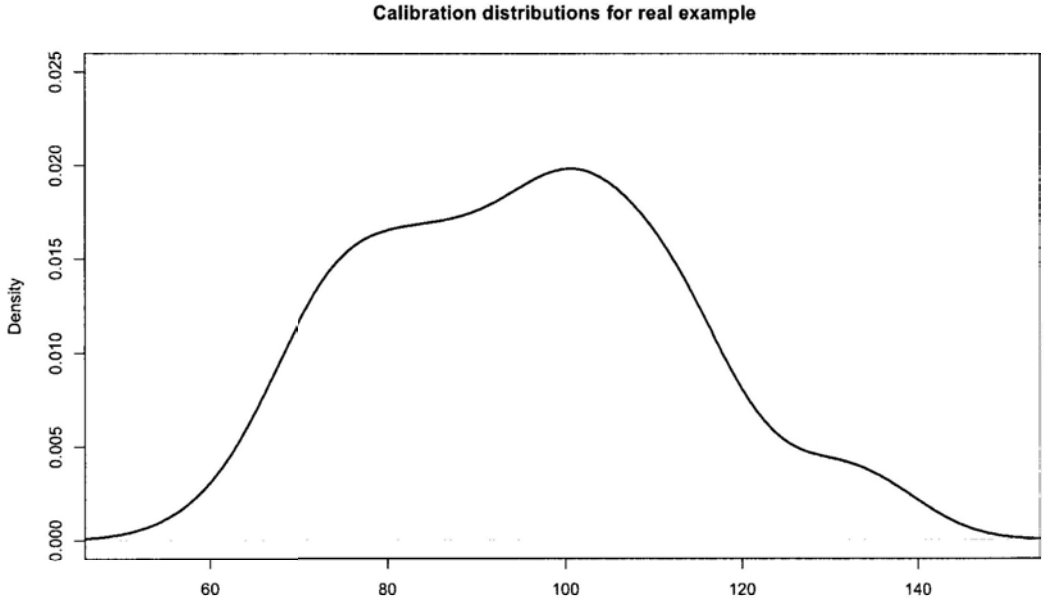


Figure 3.3: Real Example

Chapter 4

L_v Measure for Model Selection of Two-level Structural Equation Models

4.1 Introduction

In the previous chapters, L_v measure has been used for model selection of nonlinear structural equation models (SEMs) with different types of data, which are obtained from a single population. However, in many applications, the observations may exhibit two possible kinds of heterogeneity. The first kind is mixture data which means that the observations are obtained from K $(k > 1)$ populations with different distributions. In this kind of heterogenous data, K may be known or unknown and is usually small. Besides, no information is available on which of the K populations an individual observation belongs to. The second kind of heterogeneous data are drawn from a number of different groups (clusters) with a known hierarchical structure. Examples may be drawing random samples of patients from within random samples of clinics or hospitals; or of students from within random samples of schools. In contrast to the mixture data, these hierarchically structured data usually involve a large number of G groups, and the group membership of each observation can be specified accurately.

In this chapter, we discuss the second type of heterogeneous data, in which individuals within a group are allowed to share certain common influential factors. Hence, the assumption of independence among observations is violated when dealing with this kind of data. Clearly, ignoring the correlated structure of the data and analyzing them as observations from a single random sample give erroneous results. Moreover, it is also desirable to establish a meaningful model for the between-groups levels, and study the effects of the between-groups latent variables to the within-groups latent variables. Consequently, two-level SEMs that take into consideration of the correlated structure of the data are developed. Recently, many statistical methods have been proposed for the analysis of this model. For example, maximum likelihood (ML) approach (see McDonald and Goldstein, 1989; Zhang and Lee, 2001; Lee and Shi, 2001; Lee and Song, 2005) and Bayesian approach (see Song and Lee, 2004; Lee, 2007). In this chapter, we focus on model selection which is also an important issue in data analysis. Song and Lee (2004) and Lee (2007, chap. 9) applied Bayes factor to model selection in two-level SEMs. In this chapter, the L_v measure (Ibrahim et al., 2001; Chen et al., 2004) is applied to model selection of two-level nonlinear SEMs. According to the discussions in previous chapters, L_v measure can be viewed as a Bayesian goodness-of-fit statistics. Through MCMC methods, the L_v measure can be computed easily after obtaining the estimates of unknown parameters and latent variables. In addition to applying the L_v measure to two-level nonlinear SEMs, the calibration distribution is defined and used for model selection.

The remainder of this chapter is organized as follows. A two-level nonlinear SEM is defined in Section 4.2. L_v measure for model selection in two-level nonlinear SEMs is discussed in Section 4.3. In Section 4.4, a simulation study is conducted to evaluate the performance of the L_v measure for model selection. In Section 4.5, a real example is analyzed through the proposed methodology. Finally, a discussion is presented in Section 4.6.

4.2 Model Description

Suppose $\{\boldsymbol{y}_{gi}, i = 1, \cdots, N_g, g = 1, \cdots, G\}$ is a collection of p-variate random vectors. The sample size N_g may differ from group to group so that the data set is unbalanced. At the first level, we assume that, conditional on the group mean $\boldsymbol{V}_g$, random observations in each group satisfy the following measurement equation:

$$\boldsymbol{y}_{gi} = \boldsymbol{V}_g + \boldsymbol{\Lambda}_{1g}\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}, \tag{4.1}$$

where $\boldsymbol{\omega}_{1gi}$ is a $q_1 \times 1$ random vector of latent variables; $\boldsymbol{\Lambda}_{1g}$ is a $p \times q_1$ loading matrix; $\boldsymbol{\epsilon}_{1gi}$ is a $p \times 1$ random vector of error measurements which is independent of $\boldsymbol{\omega}_{1gi}$, and we assume that $\boldsymbol{\epsilon}_{1gi} \sim N[\boldsymbol{0}, \boldsymbol{\Psi}_{1g}]$, where $\boldsymbol{\Psi}_{1g}$ is a diagonal matrix. Note that due to the existence of $\boldsymbol{V}_g$, $\boldsymbol{y}_{gi}$ and $\boldsymbol{y}_{gj}$ are not independent. Hence, in the two-level SEM, the usual assumption on the independence of the observations is violated. To deal with this difficulty, an equation for the between-groups level is considered: for $g = 1, \cdots, G$,

$$\boldsymbol{V}_g = \boldsymbol{u} + \boldsymbol{\Lambda}_2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g}, \tag{4.2}$$

where $\boldsymbol{u}$ is the vector of intercepts; $\boldsymbol{\Lambda}_2$ is a $p \times q_2$ loading matrix; $\boldsymbol{\omega}_{2g}$ is a $q_2 \times 1$ vector of latent variables; $\boldsymbol{\epsilon}_{2g}$ is a $p \times 1$ random vector of error measurements which is independent of $\boldsymbol{\omega}_{2g}$, and we assume that $\boldsymbol{\epsilon}_{2g} \sim N[\boldsymbol{0}, \boldsymbol{\Psi}_2]$, where $\boldsymbol{\Psi}_2$ is a diagonal matrix. Moreover, the first and the second level measurement errors are assumed to be independent. According to equations (4.1) and (4.2), $\boldsymbol{y}_{gi}$ can be expressed as

$$\boldsymbol{y}_{gi} = \boldsymbol{u} + \boldsymbol{\Lambda}_2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g} + \boldsymbol{\Lambda}_{1g}\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}. \tag{4.3}$$

To assess the inter-relationships among the latent variables, $\boldsymbol{\omega}_{igi}$ and $\boldsymbol{\omega}_{2g}$ are partitioned as $\boldsymbol{\omega}_{1gi} = (\boldsymbol{\eta}_{1gi}^T, \boldsymbol{\xi}_{1gi}^T)^T$ and $\boldsymbol{\omega}_{2g} = (\boldsymbol{\eta}_{2g}^T, \boldsymbol{\xi}_{2g}^T)^T$, respectively, where $\boldsymbol{\eta}_{1gi}(q_{11} \times 1)$ and $\boldsymbol{\eta}_{2g}(q_{21} \times 1)$ are endogenous latent variables, $\boldsymbol{\xi}_{1gi}(q_{12} \times 1)$ and $\boldsymbol{\xi}_{2g}(q_{22} \times 1)$ are exogenous latent variables, $q_{j1} + q_{j2} = q_j$ for $j = 1, 2$. Furthermore, we assume that $\boldsymbol{\xi}_{1gi} \sim N[\boldsymbol{0}, \boldsymbol{\Phi}_{1g}]$ and $\boldsymbol{\xi}_{2g} \sim N[\boldsymbol{0}, \boldsymbol{\Phi}_2]$. The

following nonlinear structural equations are incorporated in the between-groups and within-groups models of the proposed two-level SEM:

$$\boldsymbol{\eta}_{1gi} = \boldsymbol{\Pi}_{1g}\boldsymbol{\eta}_{1gi} + \boldsymbol{\Gamma}_{1g}\boldsymbol{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi}, \tag{4.4}$$

$$\boldsymbol{\eta}_{2g} = \boldsymbol{\Pi}_{2}\boldsymbol{\eta}_{2g} + \boldsymbol{\Gamma}_{2}\boldsymbol{F}_2(\boldsymbol{\xi}_{2g}) + \boldsymbol{\delta}_{2g}, \tag{4.5}$$

where $\boldsymbol{F}_1(\boldsymbol{\xi}_{1gi}) = (f_{11}(\boldsymbol{\xi}_{1gi}), \cdots, f_{1a}(\boldsymbol{\xi}_{1gi}))^T$ and $\boldsymbol{F}_2(\boldsymbol{\xi}_{2g}) = (f_{21}(\boldsymbol{\xi}_{2g}), \cdots, f_{2b}(\boldsymbol{\xi}_{2g}))^T$ are vector-valued functions with nonzeros differentiable functions f_{1k} and f_{2k}, respectively, and usually $a \geq q_{12}$ and $b \geq q_{22}$; $\boldsymbol{\Pi}_{1g}(q_{11} \times q_{11}), \boldsymbol{\Pi}_2(q_{21} \times q_{21})$, $\boldsymbol{\Gamma}_{1g}(q_{11} \times a)$, and $\boldsymbol{\Gamma}_2(q_{21} \times b)$ are unknown parameter matrices; $\boldsymbol{\delta}_{1gi}$ and $\boldsymbol{\delta}_{2g}$ are vectors of error measurements, and we assume that $\boldsymbol{\delta}_{1gi} \sim N[\mathbf{0}, \boldsymbol{\Psi}_{1g\delta}]$ and $\boldsymbol{\delta}_{2g} \sim N[\mathbf{0}, \boldsymbol{\Psi}_{2\delta}]$, where $\boldsymbol{\Psi}_{1g\delta}$ and $\boldsymbol{\Psi}_{2\delta}$ are diagonal matrices. In the within-groups structural equation and between-groups structural equation, we assume that $\boldsymbol{\delta}_{1gi}$ and $\boldsymbol{\delta}_{2g}$ are independent of $\boldsymbol{\xi}_{1gi}$ and $\boldsymbol{\xi}_{2g}$, respectively. Let $\boldsymbol{\Lambda}^*_{1g} = (\boldsymbol{\Pi}_{1g}, \boldsymbol{\Gamma}_{1g})$, $\boldsymbol{G}_1(\boldsymbol{\omega}_{1gi}) = (\boldsymbol{\eta}^T_{1gi}, \boldsymbol{F}_1(\boldsymbol{\xi}_{1gi})^T)^T$, $\boldsymbol{\Lambda}^*_2 = (\boldsymbol{\Pi}_1, \boldsymbol{\Gamma}_2)$, and $\boldsymbol{G}_2(\boldsymbol{\omega}_{2g}) = (\boldsymbol{\eta}^T_{2g}, \boldsymbol{F}_2(\boldsymbol{\xi}_{2g})^T)^T$, equations (4.4) and (4.5) can be rewritten as:

$$\boldsymbol{\eta}_{1gi} = \boldsymbol{\Lambda}^*_{1g}\boldsymbol{G}_1(\boldsymbol{\omega}_{1gi}) + \boldsymbol{\delta}_{1gi}, \tag{4.6}$$

$$\boldsymbol{\eta}_{2g} = \boldsymbol{\Lambda}^*_{2}\boldsymbol{G}_2(\boldsymbol{\omega}_{2g}) + \boldsymbol{\delta}_{2g}. \tag{4.7}$$

Moreover, we assume that the within-groups latent vectors $\boldsymbol{\eta}_{1gi}$ and $\boldsymbol{\xi}_{1gi}$ are independent of the between-groups latent vectors $\boldsymbol{\eta}_{2g}$ and $\boldsymbol{\xi}_{2g}$. Thus, this two-level nonlinear SEM dose not accommodate the effects of the latent variables in the between-groups level on the latent variables in the within-groups level. However, in the within-groups model or in the between-groups model, nonlinear effects of the exogenous latent variables on the endogenous latent variables can be assessed through (4.4) and (4.5); and the hierarchical structure of the data has been taken into account. As the functions f_{1k} in $\boldsymbol{F}_1(\boldsymbol{\xi}_{1gi})$ and f_{2k} in $\boldsymbol{F}_2(\boldsymbol{\xi}_{2g})$ are rather general, the common interaction and quadratic effects are their special cases. Practically, allowing nonlinear relationships such as interaction and quadratic terms among latent variables leads to models that represent reality more accurately.

Furthermore, we assume that $\boldsymbol{\Pi}_{1g0} = \boldsymbol{I}_1 - \boldsymbol{\Pi}_{1g}$ and $\boldsymbol{\Pi}_{20} = \boldsymbol{I}_2 - \boldsymbol{\Pi}_2$ are nonsingular, and their determinants are respectively independent of the elements in $\boldsymbol{\Pi}_{1g}$ and $\boldsymbol{\Pi}_2$, $\boldsymbol{\Lambda}_{1g\eta}$ and $\boldsymbol{\Lambda}_{1g\xi}$ are submatrices of $\boldsymbol{\Lambda}_{1g}$ corresponding to $\boldsymbol{\eta}_{1gi}$ and $\boldsymbol{\xi}_{1gi}$, respectively; and $\boldsymbol{\Lambda}_{2\eta}$ and $\boldsymbol{\Lambda}_{2\xi}$ are submatrices of $\boldsymbol{\Lambda}_2$ corresponding to $\boldsymbol{\eta}_{2g}$ and $\boldsymbol{\xi}_{2g}$, respectively. Then the two-level nonlinear SEM defined by equations (4.1), (4.2), (4.4), and (4.5) can be rewritten as

$$\begin{aligned} \boldsymbol{y}_{gi} = \boldsymbol{u} &+ \boldsymbol{\Lambda}_{2\eta}\left(\boldsymbol{\Pi}_{20}^{-1}\boldsymbol{\Gamma}_2\boldsymbol{F}_2(\boldsymbol{\xi}_{2g}) + \boldsymbol{\delta}_{2g}\right) + \boldsymbol{\Lambda}_{2\xi}\boldsymbol{\xi}_{2g} + \boldsymbol{\epsilon}_{2g} \\ &+ \boldsymbol{\Lambda}_{1g\eta}\left(\boldsymbol{\Pi}_{1g0}^{-1}\boldsymbol{\Gamma}_1\boldsymbol{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi}\right) + \boldsymbol{\Lambda}_{1g\xi}\boldsymbol{\xi}_{1gi} + \boldsymbol{\epsilon}_{1gi}. \end{aligned} \tag{4.8}$$

The proposed two-level nonlinear SEM is not identified without imposing the identification restrictions. The common method of fixing appropriate elements in $\boldsymbol{\Lambda}_{1g}, \boldsymbol{\Pi}_{1g}, \boldsymbol{\Gamma}_{1g}, \boldsymbol{\Lambda}_2, \boldsymbol{\Pi}_2$, and $\boldsymbol{\Gamma}_2$ at preassigned values is used to achieve an identified model. For convenience, we denote the proposed two-level SEM by M, and let $\boldsymbol{\theta}$ be the parameter vector that contains all the unknown structural parameters in $\boldsymbol{u}$, $\boldsymbol{\Lambda}_{1g}$, $\boldsymbol{\Psi}_{1g}$, $\boldsymbol{\Pi}_{1g}$, $\boldsymbol{\Gamma}_{1g}$, $\boldsymbol{\Phi}_{1g}$, $\boldsymbol{\Psi}_{1g\delta}$, $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Psi}_2$, $\boldsymbol{\Gamma}_2$, and $\boldsymbol{\Phi}_2$. Let $\boldsymbol{Y}^{obs} = \{\boldsymbol{y}_{gi}, g = 1, \cdots, G, i = 1, \cdots, N_g\}$ denote the observed data, and $\boldsymbol{Y}^{rep} = \{\boldsymbol{y}_{gi}^{rep}, g = 1, \cdots, G, i = 1, \cdots, N_g\}$, which has the same distribution with $\boldsymbol{Y}^{obs}$, denote the replicated data. Let $\boldsymbol{V} = (\boldsymbol{V}_1, \cdots, \boldsymbol{V}_G)$, $\boldsymbol{\Omega}_{1g} = (\boldsymbol{\omega}_{1g1}, \cdots, \boldsymbol{\omega}_{1gN_g})$, $\boldsymbol{\Omega}_1 = (\boldsymbol{\Omega}_{11}, \cdots, \boldsymbol{\Omega}_{1G})$, and $\boldsymbol{\Omega}_2 = (\boldsymbol{\omega}_{21}, \cdots, \boldsymbol{\omega}_{2G})$. Moreover, let $\boldsymbol{\Omega}_{1\eta}$ and $\boldsymbol{\Omega}_{1\xi}$ be the sub-matrices of $\boldsymbol{\Omega}_1$ corresponding to $\boldsymbol{\eta}_{1gi}$ and $\boldsymbol{\xi}_{1gi}$, respectively, and let $\boldsymbol{\Omega}_{2\eta}$ and $\boldsymbol{\Omega}_{2\xi}$ be the sub-matrices of $\boldsymbol{\Omega}_2$ corresponding to $\boldsymbol{\eta}_{2g}$ and $\boldsymbol{\xi}_{2g}$, respectively. Furthermore, let Θ be the space of the parameter vector $\boldsymbol{\theta}$, Ξ_1 be the space of the latent variables $\boldsymbol{\xi}_{1gi}$ for $g = 1, \cdots, G$ and $i = 1, \cdots, N_g$, and let Ξ_2 be the space of the latent variables $\boldsymbol{\xi}_{2g}$ for $g = 1, \cdots, G$.

4.3 L_v Measure for Two-level Structural Equation Models

4.3.1 Definition of the L_v Measure

According to the definition of the L_v measure defined in previous chapters, we define the L_v measure for the proposed two-level SEM as

$$\begin{aligned} L_v(\boldsymbol{Y}^{obs}, M) = & \sum_{g=1}^{G}\sum_{i=1}^{N_g} \mathrm{tr}\left\{\mathrm{Var}(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{Y}^{obs}, M)\right\} \\ & + v\sum_{g=1}^{G}\sum_{i=1}^{N_g} \mathrm{tr}\left\{(\boldsymbol{\mu}_{gi} - \boldsymbol{y}_{gi})(\boldsymbol{\mu}_{gi} - \boldsymbol{y}_{gi})^T\right\}, \end{aligned} \tag{4.9}$$

where $v \in [0, 1]$, and $\boldsymbol{\mu}_{gi} = E(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{Y}^{obs}, M)$, in which the conditional expectation is taken with respect to the posterior predictive distribution:

$$\begin{aligned} & p(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{Y}^{obs}, M) \\ & = \int_{\Theta\times\Xi_1\times\Xi_2} p(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}, M)p(\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}|\boldsymbol{Y}^{obs}, M)d\boldsymbol{\theta}d\boldsymbol{\xi}_{1gi}d\boldsymbol{\xi}_{2g}. \end{aligned} \tag{4.10}$$

Therefore,

$$\boldsymbol{\mu}_{gi} = E\left[E(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g})|\boldsymbol{Y}^{obs}, M\right]. \tag{4.11}$$

According to the definition of the proposed model, the conditional expectation $E(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}, M)$ is given by

$$\begin{aligned} E(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}, M) = & \boldsymbol{u} + \boldsymbol{\Lambda}_{2\eta}(\boldsymbol{\Pi}_{20}^{-1}\boldsymbol{\Gamma}_2\boldsymbol{F}_2(\boldsymbol{\xi}_{2g})) + \boldsymbol{\Lambda}_{2\xi}\boldsymbol{\xi}_{2g} \\ & + \boldsymbol{\Lambda}_{1g\eta}(\boldsymbol{\Pi}_{1g0}^{-1}\boldsymbol{\Gamma}_{1g}\boldsymbol{F}_1(\boldsymbol{\xi}_{1gi})) + \boldsymbol{\Lambda}_{1g\xi}\boldsymbol{\xi}_{1gi}. \end{aligned} \tag{4.12}$$

Similarly, the conditional variance $\mathrm{Var}(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{Y}^{obs})$ in equation (4.9) is given by

$$\begin{aligned} \mathrm{Var}(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{Y}^{obs}, M) = & \mathrm{Var}\left[E(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}, M)|\boldsymbol{Y}^{obs}, M\right] \\ & + E\left[\mathrm{Var}(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}, M)|\boldsymbol{Y}^{obs}, M\right]. \end{aligned} \tag{4.13}$$

Since $\boldsymbol{\eta}_{1gi}$ and $\boldsymbol{\eta}_{2g}$ are independent, we get

$$\begin{aligned}\mathrm{Var}(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}, M) &= (\boldsymbol{\Lambda}_{2\eta}\boldsymbol{\Pi}_{20}^{-1})\boldsymbol{\Psi}_{2\delta}(\boldsymbol{\Lambda}_{2\eta}\boldsymbol{\Pi}_{20}^{-1})^T + \boldsymbol{\Psi}_2 \\ &\quad + (\boldsymbol{\Lambda}_{1g\eta}\boldsymbol{\Pi}_{1g0}^{-1})\boldsymbol{\Psi}_{1g\delta}(\boldsymbol{\Lambda}_{1g\eta}\boldsymbol{\Pi}_{1g0}^{-1})^T + \boldsymbol{\Psi}_{1g}.\end{aligned} \tag{4.14}$$

The first term in equation (4.13) is

$$\begin{aligned}&\mathrm{Var}\left[E(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}, M)|\boldsymbol{Y}^{obs}, M\right] \\ &= E\left[E(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}, M)\left(E(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2g}, M)\right)^T|\boldsymbol{Y}^{obs}, M\right] - \boldsymbol{\mu}_{gi}\boldsymbol{\mu}_{gi}^T.\end{aligned} \tag{4.15}$$

Based on the above discussion, the L_v measure can be calculated easily after obtaining the conditional expectation given by equation (4.11) and the conditional variance given by equation (4.13). However, these two terms cannot be obtained directly because the high dimensional integrals involved in the conditional distribution of $p(\boldsymbol{\theta}, \boldsymbol{\xi}_{1gi}, \boldsymbol{\xi}_{2gi}|\boldsymbol{Y}^{obs}, M)$. Hence, the closed form of the L_v measure cannot be obtained, and MCMC methods are applied to calculate the L_v measure. The model with the smallest value of the L_v measure is selected.

4.3.2 Computation of the L_v Measure

With observations from the posterior simulation via MCMC methods, the L_v measure can be calculated. Our strategy for the computation of the L_v measure is to augment the observed data with the latent data that come from the latent variables and/or latent measurements, then MCMC tools are applied to simulate observations in the posterior analysis. More specifically, we consider the joint posterior distribution of $[\boldsymbol{\theta}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{V}|\boldsymbol{Y}^{obs}, M]$. The Gibbs sampler (Geman and Geman, 1984) is used to generate a sequence of observations from this joint posterior distribution. Based on these observations, the Bayesian estimation of the unknown parameters can be obtained, and then the L_v measure can be computed easily. In applying the Gibbs sampler, we iteratively sample from the following conditional distributions: $p(\boldsymbol{V}|\boldsymbol{\theta}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{Y}^{obs}, M)$, $p(\boldsymbol{\Omega}_1|\boldsymbol{\theta}, \boldsymbol{\Omega}_2, \boldsymbol{V}, \boldsymbol{Y}^{obs}, M)$, $p(\boldsymbol{\Omega}_2|\boldsymbol{\theta}, \boldsymbol{\Omega}_1, \boldsymbol{V},$

$\boldsymbol{Y}^{obs}, M)$, and $p(\boldsymbol{\theta}|\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{V}, \boldsymbol{Y}^{obs}, M)$. After the convergence of the MCMC algorithm, a Gibbs sample $\{\boldsymbol{\theta}^{(r)}, \boldsymbol{\Omega}_1^{(r)}, \boldsymbol{\Omega}_2^{(r)}, \boldsymbol{V}^{(r)}, r = 1, \cdots, R\}$ can be recorded, and with this sample, the estimation of the unknown parameters and latent variables are given as follows:

$$\hat{\boldsymbol{\theta}} = \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\theta}^{(r)},\ \hat{\boldsymbol{\omega}}_{2g} = \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\omega}_{2g}^{(r)},\ \hat{\boldsymbol{\omega}}_{1gi} = \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\omega}_{1gi}^{(r)}.$$

Furthermore, let

$$\begin{aligned}\boldsymbol{\mu}_{gi}^{(r)} &= \boldsymbol{u}^{(r)} + \boldsymbol{\Lambda}_{2\eta}^{(r)}\{(\boldsymbol{\Pi}_{20}^{(r)})^{-1}\boldsymbol{\Gamma}_2^{(r)}\boldsymbol{F}_2(\boldsymbol{\xi}_{2g}^{(r)})\} + \boldsymbol{\Lambda}_{2\xi}^{(r)}\xi_{2g}^{(r)} \\ &\quad + \boldsymbol{\Lambda}_{1g\eta}^{(r)}\{(\boldsymbol{\Pi}_{1g0}^{(r)})^{-1}\boldsymbol{\Gamma}_{1g}^{(r)}\boldsymbol{F}_1(\boldsymbol{\xi}_{1gi}^{(r)})\} + \boldsymbol{\Lambda}_{1g\xi}^{(r)}\xi_{1gi}^{(r)},\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{gi}^{(r)} &= \{\boldsymbol{\Lambda}_{2\eta}^{(r)}(\boldsymbol{\Pi}_{20}^{(r)})^{-1}\}\boldsymbol{\Psi}_{2\delta}^{(r)}\{\boldsymbol{\Lambda}_{2\eta}^{(r)}(\boldsymbol{\Pi}_{20}^{(r)})^{-1}\}^T + \boldsymbol{\Psi}_2^{(r)} \\ &\quad + \{\boldsymbol{\Lambda}_{1g\eta}^{(r)}(\boldsymbol{\Pi}_{1g0}^{(r)})^{-1}\}\boldsymbol{\Psi}_{1g\delta}^{(r)}\{\boldsymbol{\Lambda}_{1g\eta}^{(r)}(\boldsymbol{\Pi}_{1g0}^{(r)})^{-1}\}^T + \boldsymbol{\Psi}_{1g}^{(r)},\end{aligned}$$

and

$$\hat{\boldsymbol{\mu}}_{gi} = \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\mu}_{gi}^{(r)},\quad \widehat{Var}(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{Y}^{obs}) = \frac{1}{R}\sum_{r=1}^{R}\left[\boldsymbol{\Sigma}_{gi}^{(r)} + \boldsymbol{\mu}_{gi}^{(r)}(\boldsymbol{\mu}_{gi}^{(r)})^T\right] - \hat{\boldsymbol{\mu}}_{gi}\hat{\boldsymbol{\mu}}_{gi}^T.$$

Then L_v measure is computed as follow:

$$L_v(\boldsymbol{Y}_{obs}) = \sum_{g=1}^{G}\sum_{i=1}^{N_g}\left[\mathrm{tr}\left\{\widehat{Var}(\boldsymbol{y}_{gi}^{rep}|\boldsymbol{Y}^{obs})\right\} + \mathrm{tr}\left\{v(\hat{\boldsymbol{\mu}}_{gi} - \boldsymbol{y}_{gi})(\hat{\boldsymbol{\mu}}_{gi} - \boldsymbol{y}_{gi})^T\right\}\right]. \tag{4.16}$$

4.3.3 Full Conditional Distributions

In this section, we discuss the conditional distributions required in Gibbs sampler. First, we consider the conditional distribution $p[\boldsymbol{\theta}|\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{V}, \boldsymbol{Y}^{obs}, M]$. For the proposed two-level nonlinear SEM, this conditional distribution can be further decomposed into components involving various structural parameters in the between-groups and within-groups models. These components are different under various special cases of the model. Three typical examples are:

(1). Models with different within-groups parameters across groups. In this case, the within-groups structural parameters $\boldsymbol{\theta}_{1g} = \{\boldsymbol{\Lambda}_{1g}, \boldsymbol{\Psi}_{1g}, \boldsymbol{\Pi}_{1g}, \boldsymbol{\Gamma}_{1g}, \boldsymbol{\Phi}_{1g}, \boldsymbol{\Psi}_{1g\delta}\}$. Practically, G and N_g should not be too small for drawing valid statistical conclusions for the between-groups model and the g-th within-groups model.

(2). Models with some invariant within-groups parameters. In this case, parameters involved in $\boldsymbol{\theta}_{1g}$ associated with the g-th group are equal to those associated with some other groups.

(3). Models with all invariant within-groups parameters. Under this situation, $\boldsymbol{\theta}_{11} = \boldsymbol{\theta}_{12} = \cdots = \boldsymbol{\theta}_{1G}$.

Conditional distributions under these three cases are similar but different. We take the first case for an illustration. To get these conditional distributions, prior distributions of the parameters are required. On the basis of the reasons given in the Bayesian literature of SEMs, conjugate type prior distributions are used. The prior distributions for parameters involved in the within-groups model are given as follows:

$$p(\boldsymbol{\Phi}_{1g}^{-1}) \triangleq W_{q_{12}}[\boldsymbol{R}_{01g}, \rho_{01g}],$$

for $k = 1, \cdots, p$,

$$P(\psi_{1gk}^{-1}) \triangleq Gamma[\alpha_{01gk}, \beta_{01gk}], \quad p(\boldsymbol{\Lambda}_{1gk}|\psi_{1gk}) \triangleq N[\boldsymbol{\Lambda}_{01gk}, \psi_{1gk}\boldsymbol{H}_{01gk}],$$

and for $k = 1, \cdots, q_{11}$,

$$P(\psi_{1g\delta k}^{-1}) \triangleq Gamma[\alpha_{01g\delta k}, \beta_{01g\delta k}], \quad p(\boldsymbol{\Lambda}_{1gk}^{*}|\psi_{1gk}) \triangleq N[\boldsymbol{\Lambda}_{01gk}^{*}, \psi_{1gk}\boldsymbol{H}_{01gk}^{*}],$$

where $\boldsymbol{\Lambda}_{1gk}$ and $\boldsymbol{\Lambda}_{1gk}^{*}$ are the kth rows of $\boldsymbol{\Lambda}_{1g}$ and $\boldsymbol{\Lambda}_{1g}^{*}$, respectively; ψ_{1gk} and $\psi_{1g\delta k}$ are the kth diagonal elements of $\boldsymbol{\Psi}_{1g}$ and $\boldsymbol{\Psi}_{1g\delta}$, respectively; $\rho_{01g}, \alpha_{01gk}, \beta_{01gk}, \boldsymbol{\Lambda}_{01gk}, \alpha_{01g\delta k}, \beta_{01g\delta k}, \boldsymbol{\Lambda}_{01gk}^{*}$, and the positive definite matrices $\boldsymbol{R}_{01g}, \boldsymbol{H}_{01gk}$, and $\boldsymbol{H}_{01gk}^{*}$ are all given hyperparameters.

For the parameters involved in the between-groups model, the following prior distributions are considered:

$$p(\boldsymbol{u}) \triangleq N[\boldsymbol{u}_0, \Sigma_0], \; p(\boldsymbol{\Phi}_2^{-1}) \triangleq W_{q_{22}}[\boldsymbol{R}_{02}, \rho_{02}],$$

for $k = 1, \cdots, p$,

$$P(\psi_{2k}^{-1}) \triangleq Gamma[\alpha_{02k}, \beta_{02k}],\ p(\boldsymbol{\Lambda}_{2k}|\psi_{2k}) \triangleq N[\boldsymbol{\Lambda}_{02k}, \psi_{2k}\boldsymbol{H}_{02k}],$$

and for $k = 1, \cdots, q_{21}$,

$$P(\psi_{2\delta k}^{-1}) \triangleq Gamma[\alpha_{02\delta k}, \beta_{02\delta k}],\ p(\boldsymbol{\Lambda}_{2k}^{*}|\psi_{2k}) \triangleq N[\boldsymbol{\Lambda}_{02k}^{*}, \psi_{2k}\boldsymbol{H}_{02k}^{*}],$$

where $\boldsymbol{\Lambda}_{2k}$ and $\boldsymbol{\Lambda}_{2k}^{*}$ are the kth rows of $\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Lambda}_2^{*}$, respectively; ψ_{2k} and $\psi_{2\delta k}$ are the kth diagonal elements of $\boldsymbol{\Psi}_2$ and $\boldsymbol{\Psi}_{2\delta}$, respectively; $\boldsymbol{u}_0$, ρ_{02}, α_{02k}, β_{02k}, $\alpha_{02\delta k}$, $\beta_{02\delta k}$, $\boldsymbol{\Lambda}_{02k}$, $\boldsymbol{\Lambda}_{02k}^{*}$, and the positive definite matrices $\boldsymbol{\Sigma}_0$, $\boldsymbol{R}_{02}$, H_{02k}, and $\boldsymbol{H}_{02k}^{*}$ are all given hyperparameters.

Given the above conjugate prior distributions, the conditional distributions of the components in $p(\boldsymbol{\theta}|\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{V}, \boldsymbol{Y}^{obs})$ can be easily obtained. These conditional distributions are generalizations of those that are associated with a single level model; and most of them are standard distributions, such as normal, univariate truncated normal, Gamma, and inverted Wishart distributions. Simulating observations from them requires little computing time. The posterior density functions of $p(\boldsymbol{V}|\boldsymbol{\theta}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{Y}^{obs}, M)$, $p(\boldsymbol{\Omega}_1|\boldsymbol{\theta}, \boldsymbol{\Omega}_2, \boldsymbol{V}, \boldsymbol{Y}^{obs}, M)$, and $p(\boldsymbol{\Omega}_2|\boldsymbol{\theta}, \boldsymbol{\Omega}_1, \boldsymbol{V}, \boldsymbol{Y}^{obs}, M)$ can also be easily obtained. However, the conditional distributions, $p(\boldsymbol{\Omega}_1|\boldsymbol{\theta}, \boldsymbol{\Omega}_2, \boldsymbol{V}, \boldsymbol{Y}^{obs}, M)$ and $p(\boldsymbol{\Omega}_2|\boldsymbol{\theta}, \boldsymbol{\Omega}_1, \boldsymbol{V}, \boldsymbol{Y}^{obs}, M)$, are complex. Hence, it is necessary to implement the MH algorithm for an efficient simulation of observations from these conditional distributions. The conditional distributions and the MH algorithm involved in this chapter can be derived and implemented similarly as before. They are not presented to save space.

4.3.4 Calibration Distribution

As discussed in previous chapters, criterion-based methods typically rely on the minimum criterion value as the basis for model selection. However this basis is not satisfactory in general, since it may be misleading when two different models have similar values of the L_v measure. Thus, one of

the crucial steps in using criterion-based method for model assessment and model choice is to define a calibration for the criterion. In this chapter, we apply the same calibration as discussed in previous chapters to the implementation of the L_v measure. Specifically, let $L_v(\boldsymbol{Y}^{obs}, M_c)$ denote the L_v measure of the candidate model M_c, and $L_v(\boldsymbol{Y}^{obs}, M_t)$ denote the L_v measure of the true model M_t. Given v, the difference of the L_v measures between the candidate model M_c and the true model M_t is defined as

$$D_v(\boldsymbol{Y}^{obs}, M_c) \equiv L_v(\boldsymbol{Y}^{obs}, M_c) - L_v(\boldsymbol{Y}^{obs}, M_t). \tag{4.17}$$

Then the marginal distribution of $D_v(\boldsymbol{Y}^{obs}, M_c)$, computed with respect to the prior predictive distribution of $\boldsymbol{Y}^{obs}$ under the true model M_t, is defined as the calibration distribution. The prior predictive distribution for the proposed model is defined by

$$p_t(\boldsymbol{Y}^{obs}) = \int_{\Theta \times \Xi_1 \times \Xi_2} p(\boldsymbol{Y}^{obs}|\boldsymbol{\theta}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, M_t) p(\boldsymbol{\theta}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2|M_t) d\boldsymbol{\theta} d\boldsymbol{\Omega}_1 d\boldsymbol{\Omega}_2. \tag{4.18}$$

We denote the calibration distribution of the candidate model M_c by

$$PL_c \equiv p(D_v(\boldsymbol{Y}^{obs}, M_c)). \tag{4.19}$$

Similarly, several statistical summaries of calibration distribution are considered. These include highest probability density (HPD) interval, the mean $\mu_v(M_c)$, and the standard deviation $SD_v(M_c)$ of $D_v(\boldsymbol{Y}^{obs}, M_c)$. HPD interval denotes the shortest credible interval which means the interval with the highest posterior density. $\mu_v(M_c)$ measures, on average, how close the candidate model and true model are. $SD_v(M_c)$ measures the dispersion of the calibration distribution. Specifically, if the candidate model is "close" to the true model, $\mu_v(M_c)$ is close to zero; otherwise, $\mu_v(M_c)$ is far apart from zero. From equation (4.19), to well define PL_c, proper prior distributions for the unknown parameters in $\boldsymbol{\theta}$ are needed. The definition given by equation (4.19) is appealing since it avoids the potential problem of a double use of the data (see Ibrahim et al., 2001).

For the proposed two-level nonlinear SEM, we can get a closed form of neither the L_v measure, nor the calibration distribution. So MCMC methods are used to estimate the calibration distributions of the considered competing models. Specifically, for a candidate model M_c, a sample of $D_v(\boldsymbol{Y}^{obs}, M_c)$ is generated via Gibbs sampler algorithm, then kernel density estimation method (Silverman, 1986; Sheather and Jones, 1991; Scott, 1992) is used to estimate the distribution of $D_v(\boldsymbol{Y}^{obs}, M_c)$. However, in data analysis, true model is unknown. Therefore, the model with the smallest L_v measure is considered as the true model M_t, and a pseudo data set $\tilde{\boldsymbol{Y}}$ can be generated from this model. Details about the estimation of the calibration distributions are given in Chapter 2.

4.4 A Simulation Study

L_v measure for model selection

In this simulation study, observations are generated from a two-level nonlinear SEM, M_0, with measurement equations given by equations (4.1) and (4.2). In this model, we assume that $\boldsymbol{\Lambda}_{1g} = \boldsymbol{\Lambda}_1$. To identify the model, $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ of the within-groups model and between-groups model are given as

$$\boldsymbol{\Lambda}_1^T = \boldsymbol{\Lambda}_2^T = \begin{pmatrix} 1.0^* & \lambda_{21} & \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{52} & \lambda_{62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{83} & \lambda_{93} \end{pmatrix},$$

where the elements with asterisks are fixed.

The structural equations of the within-groups model and between-groups model for M_0 are given as follows:

$$\begin{aligned} \eta_{1gi} &= \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{13}\xi_{1g1}\xi_{1g2} + \delta_{1gi}, \\ \eta_{2g} &= \gamma_{21}\xi_{2g1} + \gamma_{22}\xi_{2g2} + \delta_{2g}, \end{aligned}$$

where $\boldsymbol{\omega}_{1gi} = \{\eta_{1gi}, \xi_{1gi1}, \xi_{1gi2}\}$ is the vector of latent variables in the within-groups model, and $\boldsymbol{\omega}_{2g} = \{\eta_{2g}, \xi_{2g1}, \xi_{2g2}\}$ is the vector of latent variables

in the between-groups model. The true values of some parameters are given by: $\lambda_{21} = \lambda_{31} = \lambda_{52} = \lambda_{62} = \lambda_{83} = \lambda_{93} = 0.8$, $\boldsymbol{u} = (0, \cdots, 0)^T$, $\boldsymbol{\Gamma}_{1g} = \boldsymbol{\Gamma}_1 = (\gamma_{11}, \gamma_{12}, \gamma_{13}) = (0.6, 0.6, -0.4)$, $\boldsymbol{\Gamma}_2 = (\gamma_{21}, \gamma_{22}) = (0.6, 0.6)$, $\boldsymbol{\Psi}_{1g} = \boldsymbol{\Psi}_1 = 0.64\boldsymbol{I}_9$, and $\boldsymbol{\Psi}_2 = 0.36\boldsymbol{I}_9$, where $\boldsymbol{I}_9$ is the 9-dimensional identity matrix. True values of parameters in the covariance matrices $\boldsymbol{\Phi}_{1g} = \boldsymbol{\Phi}_1$, $\boldsymbol{\Psi}_{1g\delta} = \psi_{1\delta}$, $\boldsymbol{\Phi}_2$ and $\boldsymbol{\Psi}_{2\delta} = \psi_{2\delta}$ are given by $\phi_{1,11} = \phi_{1,22} = 1.0$, $\phi_{1,12} = 0.3$, $\boldsymbol{\Phi}_2 = \boldsymbol{\Phi}_1$, $\psi_{1\delta} = 0.64$ and $\psi_{2\delta} = 0.36$. The structural equation of the between-groups is linear, and the structural equation of the within-groups is nonlinear with an intersection term. For each of the 100 replications, three samples with different sample sizes are generated. Specifically, the first sample is generated with $N = 1000, G = 150, N_1 = \cdots = N_{100} = 5, N_{101} = \cdots = N_{150} = 10$; the second sample is generated with $N = 1500, G = 200, N_1 = \cdots = N_{100} = 5, N_{101} = \cdots = N_{200} = 10$; and the third sample is generated with $N = 2000, G = 300, N_1 = \cdots = N_{200} = 5, N_{201} = \cdots = N_{300} = 10$.

To give a sensitivity analysis about the prior inputs of the hyperparameters in the prior distributions, two types of prior inputs given below are considered.

Prior I: The mean vectors, $\boldsymbol{u}_0, \boldsymbol{\Lambda}_{01gk} = \boldsymbol{\Lambda}_{01k}, \boldsymbol{\Lambda}^*_{01gk} = \boldsymbol{\Lambda}_{01k}$, and $\boldsymbol{\Lambda}_{02k}$, are taken as the true values of the corresponding parameters, the covariance matrices, $\boldsymbol{\Sigma}_0, \boldsymbol{H}_{01gk} = \boldsymbol{H}_{01k}, \boldsymbol{H}^*_{01gk} = \boldsymbol{H}^*_{01k}$, and $\boldsymbol{H}_{02k}$, are equal to identity matrices of appropriate orders. $\alpha_{01gk} = \alpha_{01k} = \alpha_{01g\delta k} = \alpha_{01\delta k} = \alpha_{02k} = \alpha_{02\delta k} = 15$, $\beta_{01gk} = \beta_{01k} = \beta_{01g\delta k} = \beta_{01\delta k} = 9$, and $\beta_{02k} = \beta_{02\delta k} = 5$; $\boldsymbol{R}_{01g} = \boldsymbol{R}_{01} = \boldsymbol{R}_{02} = 3.0\boldsymbol{I}_2$, and $\rho_{01g} = \rho_{01} = \rho_{02} = 6$.

Prior II: The mean vectors, $\boldsymbol{u}_0, \boldsymbol{\Lambda}_{01gk} = \boldsymbol{\Lambda}_{01k}, \boldsymbol{\Lambda}^*_{01gk} = \boldsymbol{\Lambda}_{01k}$, and $\boldsymbol{\Lambda}_{02k}$ are taken as two times of the true values of the corresponding parameters, the covariance matrices, $\boldsymbol{\Sigma}_0, \boldsymbol{H}_{01gk} = \boldsymbol{H}_{01k}, \boldsymbol{H}^*_{01gk} = \boldsymbol{H}^*_{01k}$, and $\boldsymbol{H}_{02k}$, are equal to four times of identity matrices of appropriate orders. $\alpha_{01gk} = \alpha_{01g\delta k} = \alpha_{02k} = \alpha_{02\delta k} = 15$, $\beta_{01gk} = \beta_{01g\delta k} = 19$, $\beta_{02k} = \beta_{02\delta k} = 11$, $\rho_{01g} = \rho_{01}$, and ρ_{02} are equal to four times of the values given in Prior I,

while the other hyperparameters values are the same as those given in Prior I.

Four competing models are considered. A single-level SEM denoted by M_1 is given by

$$M_1: \ \boldsymbol{y}_i = \boldsymbol{u} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i, \text{ and } \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i1}\xi_{i2} + \delta_i, \ i = 1, \cdots, N.$$

Three two-level SEMs with the same measurement equation as M_0 but different structural equations given below are considered. For $g = 1, \cdots, G$ and $i = 1, \cdots, N_g$,

$$\begin{aligned}
M_2 \ &: \ \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \delta_{1gi}, \\
&\quad \ \eta_{2g} = \gamma_{21}\xi_{2g1} + \gamma_{22}\xi_{2g2} + \gamma_{23}\xi_{2g1}\xi_{2g2} + \delta_{2g}, \\
M_3 \ &: \ \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \delta_{1gi}, \\
&\quad \ \eta_{2g} = \gamma_{21}\xi_{2g1} + \gamma_{22}\xi_{2g2} + \delta_{2g}, \\
M_4 \ &: \ \eta_{1gi} = \gamma_{11g}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{13}\xi_{1gi1}\xi_{1gi2} + \delta_{1gi}, \\
&\quad \ \eta_{2g} = \gamma_{21}\xi_{2g1} + \gamma_{22}\xi_{2g2} + \gamma_{23}\xi_{2g1}\xi_{2g2} + \delta_{2g}.
\end{aligned}$$

L_v measure for model selection

The mean values of the L_v measures and the calibration summaries are given in Table 4.1, where mean($L_{0.5}$) denotes the mean value of the L_v measure with $v = 0.5$, $\mu_{0.5}$, $SD_{0.5}$, and 95% HPD denote the mean, the standard deviation, and the 95% HPD interval of the calibration distribution of the corresponding model with $v = 0.5$, respectively. From this table, we see that the values of $\mu_{0.5}$ corresponding to M_1, M_2, and M_3 are larger than zero, and the 95% HPD intervals corresponding to these three models are far apart from zero. Hence we conclude that M_0 performs better than M_1, M_2, and M_3 under each case. However, for M_4, $\mu_{0.5}$ value is relatively close to zero, and the 95% HPD interval includes zero under each case. Therefore, we conclude that the performances of models M_0 and M_4 are similar under each case. According to the parsimonious principle, the simpler model

M_0 is selected. Besides, when sample size increases, the centers of the calibration distributions of M_1, M_2, and M_3 become further apart from zero. While the center of the calibration distribution of M_4 changes little. The estimated calibration distributions, estimated under Prior I via kernel density estimation method described in Section 3.4, are presented in Figure 4.1(a) to Figure 4.1(c). Figures corresponding to Prior II are similar and hence not presented. From these figures, the calibration distribution of M_4 is centered at zero under each case, while the others are far apart from zero. Hence we get the same conclusion as before. From the above discussion, M_0 is selected under each given type of prior inputs and sample size, thus it seems that the L_v measure is rather robust to prior inputs.

Bayes factor for model selection

In this section, the logarithm Bayes factors for comparing the above models are computed via the path sampling procedure. In applying the path sampling, an important step is to find a good linking model to link the competing models. In this simulation study, the linking model for M_0 and M_1 is given by

$$\boldsymbol{y}_{gi} = (1-t)\boldsymbol{v}_g + \boldsymbol{\Lambda}_1\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi},\ \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{13}\xi_{1gi1}\xi_{1gi2} + \delta_{1gi},$$
$$\boldsymbol{v}_g = \boldsymbol{u} + \boldsymbol{\Lambda}_2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g},\ \eta_{2g} = \gamma_{21}\xi_{2g1} + \gamma_{22}\xi_{2g2} + \delta_{2g}.$$

The linking models for M_0 and M_c, for $c = 2, 3, 4$, have the same measurement equation as M_0 but different structural equations given as follows:

$$\begin{aligned}
M_{t20}:&\eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + (1-t)\gamma_{13}\xi_{1gi1}\xi_{1gi2} + \delta_{1gi},\\
&\eta_{2g} = \gamma_{21}\xi_{2g1} + \gamma_{22}\xi_{2g2} + t\gamma_{23}\xi_{2g1}\xi_{2g2} + \delta_{2g},\\
M_{t30}:&\eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + (1-t)\gamma_{13}\xi_{1gi1}\xi_{1gi2} + \delta_{1gi},\\
&\eta_{2g} = \gamma_{21}\xi_{2g1} + \gamma_{22}\xi_{2g2} + \delta_{2g},\\
M_{t40}:&\eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{13}\xi_{1gi1}\xi_{1gi2} + \delta_{1gi},\\
&\eta_{2g} = \gamma_{21}\xi_{2g1} + \gamma_{22}\xi_{2g2} + \delta_{2g}.
\end{aligned}$$

Clearly, for $c = 1, \cdots, 4$, when $t = 0$, M_{tc0} reduces to M_0; whilst when $t = 1$,

M_{tc0} reduces to M_c. By differentiating the complete-data log-likelihood functions of these four linking models, we get

$$
\begin{aligned}
U_{t_{10}} &= -\sum_{g=1}^{G}\sum_{i=1}^{N_g} \boldsymbol{v}_g^T \boldsymbol{\Psi}_{1g}^{-1}(\boldsymbol{y}_{gi} - (1-t)\boldsymbol{v}_g - \boldsymbol{\Lambda}_{1g}\boldsymbol{\omega}_{1g}), \\
U_{t_{20}} &= -\sum_{g=1}^{G}\sum_{i=1}^{N_g} \frac{\eta_{1gi} - \gamma_{11}\xi_{1gi1} - \gamma_{12}\xi_{1gi2} - (1-t)\gamma_{13}\xi_{1gi1}\xi_{1gi2}}{\psi_{\delta 1}(\gamma_{13}\xi_{1gi1}\xi_{1gi2})^{-1}} \\
&\quad + \sum_{g=1}^{G} \frac{\eta_{2g} - \gamma_{21}\xi_{2g1} - \gamma_{22}\xi_{2g2} - t\gamma_{23}\xi_{2g1}\xi_{2g2}}{\psi_{\delta 2}(\gamma_{23}\xi_{2g1}\xi_{2g2})^{-1}}, \\
U_{t_{30}} &= -\sum_{g=1}^{G}\sum_{i=1}^{N_g} \frac{\eta_{1gi} - \gamma_{11}\xi_{1gi1} - \gamma_{12}\xi_{1gi2} - (1-t)\gamma_{13}\xi_{1gi1}\xi_{1gi2}}{\psi_{\delta 1}(\gamma_{13}\xi_{1gi1}\xi_{1gi2})^{-1}}, \\
U_{t_{40}} &= \sum_{g=1}^{G} \frac{\eta_{2g} - \gamma_{21}\xi_{2g1} - \gamma_{22}\xi_{2g2} - t\gamma_{23}\xi_{2g1}\xi_{2g2}}{\psi_{\delta 2}(\gamma_{23}\xi_{2g1}\xi_{2g2})^{-1}}.
\end{aligned}
$$

Consequently, $\log B_{10}, \cdots, \log B_{40}$ can be obtained based on the above differentiations (see Lee, 2007, chap. 5). Results for this simulation study are given in Table 4.2. According to this table, $\log B_{c0}$ for $c = 1, \cdots, 4$ are all negative. Therefore, based on the interpretation of logarithm of Bayes factor (Kass and Raftery, 1995), M_0 is selected under each case.

4.5 A Real Example

In this part, a real example discussed in Lee (2007, chap. 9) is analyzed. Here, we focus on model selection with the L_v measure. A brief introduction of the application is given. The Accelerated Schools for Quality Education (ASQE) Project is a huge project which was conducted for helping schools to achieve an internal cultural change in order to be self-reliant in attaining school-based goals in self-improvement. In this section, we focus on the particular issue about the causal relationships among the school values inventory, teachers job satisfaction, and their empowerment in identifying and solving the schools problems. Relationships among these latent variables at

the school level and the teacher level are important in the cultivation of their own and their peers skills in improving their teaching skills and practice. Based on the proposed two-level SEM that incorporates the effects of the between-groups (school level) latent variables to the within-groups (teacher level) latent variables, we can assess precise interrelationships among the latent variables in both levels. To save space, we only present our results based on analyses of the data that were obtained from September 1998 to August 1999. The data set is hierarchically structured with n = 1555 teachers nested in G = 50 schools. The data set is unbalanced with values of N_g ranged from 14 to 47. Three manifest variables (relating to questions: 'I proudly introduce my school as a worth-while working place to my friends'; 'I find that my attitude of value is close to my schools attitude of value'; and 'I can fully utilize my potentials in my school work.') y_{g1}, y_{g2}, and y_{g3} are taken as indicators for the latent factor, 'job satisfaction.' These variables are measured via a 7-point scale. For brevity, they are treated as continuous. The manifest variables y_{g4}, y_{g5}, and y_{g6} for the latent variable, 'school value inventory' are: (1) participation and collaboration, (2) collegiality, and (3) communication and consensus, which are respectively measured by the averages of seven, six, and ten items in the questionnaire. The manifest variables y_{g7}, y_{g8}, and y_{g9} for the latent factor, 'teachers empowerment' are: (1) decision making, (2) self efficacy, and (3) self autonomy, which are measured by the averages of four, four, and five items in the questionnaire. The sample means and standard deviations of the manifest variables are 4.139, 4.553, 4.487, 2.406, 3.171, 3.468, 0.534, 0.381, 0.601 and 1.371, 1.187, 1.181, 0.848, 0.763, 0.728, 0.499, 0.486, 0.490, respectively. Six two-level SEMs $(M_0, \cdots, M_5)$ with nine manifest variables and three latent variables are considered in this example. They have the same between-groups model, and the same within-groups measurement equation, but different within-groups structural equations. Specifically, a factor analysis model for the between-groups model is considered. Three latent variables, 'job satisfac-

tion, ω_{2g1}', 'schools value inventory, ω_{2g2}' and 'teachers empowerment, ω_{2g3}' are considered in this factor analysis model. These latent factors are allowed to be correlated, and $\boldsymbol{\omega}_{2g} = (\omega_{2g1}, \omega_{2g2}, \omega_{2g3}) \sim N[\mathbf{0}, \boldsymbol{\Phi}_2]$. Similar as discussed in Lee (2007), we consider the following common structure for the factor loading matrix:

$$\boldsymbol{\Lambda}_2^T = \begin{pmatrix} 1.0^* & \lambda_{2,21} & \lambda_{2,31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{2,52} & \lambda_{2,62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{2,83} & \lambda_{2,93} \end{pmatrix},$$

where the elements with asterisks are fixed parameters. For the within-groups model at the teachers level, we also use the same factor loading structure of $\boldsymbol{\Lambda}_2$ for $\boldsymbol{\Lambda}_1$ (with unknown elements denoted by $\lambda_{1,ij}$) to relate the latent factors to the manifest variables. Again, there are three latent factors, η_{gi}, ξ_{gi1}, and ξ_{gi2}, in the within-groups model. Similarly, based on the meaning of the corresponding questions, interpretations of η_{gi}, ξ_{gi1}, and ξ_{gi2}, are 'job satisfaction', 'schools value inventory', and 'teachers empowerment' that are directly related to the teachers. The variances and covariance of ξ_{gi1}, and ξ_{gi2} are given by $\phi_{1,11}, \phi_{1,22}$ and $\phi_{1,12}$, respectively. As 'job satisfaction' of the teachers is an important factor in education, it is important to investigate its relationships with the other latent factors. The within-groups structural equations for the six models are given as follows:

$$\begin{aligned}
M_0 &: \eta_{gi} = \gamma_1\xi_{gi1} + \gamma_2\xi_{gi2} + \gamma_4\omega_{2g1} + \gamma_5\omega_{2g2} + \gamma_6\omega_{2g3} + \delta_{gi}, \\
M_1 &: \eta_{gi} = \delta_{gi}, \\
M_2 &: \eta_{gi} = \gamma_1\xi_{gi1} + \gamma_2\xi_{gi2} + \delta_{gi}, \\
M_3 &: \eta_{gi} = \gamma_4\omega_{2g1} + \gamma_5\omega_{2g2} + \gamma_6\omega_{2g3} + \delta_{gi}, \\
M_4 &: \eta_{gi} = \gamma_1\xi_{gi1} + \gamma_2\xi_{gi2} + \gamma_3\xi_{gi1}\xi_{gi2} + \gamma_4\omega_{2g1} + \gamma_5\omega_{2g2} + \gamma_6\omega_{2g3} + \delta_{gi}, \\
M_5 &: \eta_{gi} = \gamma_1\xi_{gi1} + \gamma_2\xi_{gi2} + \gamma_3\xi_{gi1}\xi_{gi2} + \gamma_4\omega_{2g1} + \gamma_5\omega_{2g2} + \gamma_6\omega_{2g3} \\
&\qquad + \gamma_5\omega_{2g2}\omega_{2g3} + \delta_{gi}.
\end{aligned}$$

In this example, a Bayesian approach is used to obtain the estimates of the unknown parameters. L_v measure and Bayes factor are applied for

model selection. L_v measures for the six models are: $L_{0.5}(M_0) = 8549.997$, $L_{0.5}(M_1) = 9745.978$, $L_{0.5}(M_2) = 8500.443$, $L_{0.5}(M_3) = 12540.29$, $L_{0.5}(M_4) = 8407.755$, and $L_{0.5}(M_5) = 8471.401$. From these results, M_4 has the smallest L_v measure. Therefore M_4 is selected. To estimate the calibration distributions, 100 pseudo data sets generated from M_4 are used. The calibration summaries are given in Table 4.5, and the estimates of the calibration distributions are presented in Figure 4.3. Based on these results, M_4 is also selected. Logarithm Bayes factors are $\widehat{\log B_{01}} = 882.31$, $\widehat{\log B_{02}} = 240.59$, $\widehat{\log B_{03}} = 245.43$, $\widehat{\log B_{04}} = -4.018$, $\widehat{\log B_{05}} = -2.832$, and $\widehat{\log B_{54}} = -1.187$ (see Lee, 2007, chap. 9). Based on the values of logarithms of Bayes factor, M_4 is selected, which agree with our conclusion by using the L_v measure.

4.6 Discussion

In this chapter, we applied the L_v measure to model selection of two-level nonlinear SEMs. From the numerical studies, L_v measure and Bayes factor can achieve the same conclusion for model selection. However, compared with Bayes factor, the computation of the L_v measure is much simpler and faster. Besides, with the calibration distribution, L_v measure avoids the disadvantage of criterion-based method for model selection, that is when the difference of criterion values between two competing models is very small, we cannot decide which one is better. Moreover, the value v can vary from zero to one, which makes the L_v measure more flexible for model selection. Therefore, L_v measure is a better alternative method for model selection in the analysis of hierarchical data.

Table 4.1: Mean values of the L_v measure and calibration Summaries for Simulation Study

Prior	Sample size	Model	mean($L_{0.5}$)	$\mu(D_{0.5})$	$SD(D_{0.5})$	95% HPD
Prior I	N=1000	M_0	18037.049	-	-	-
		M_1	20379.092	2342.043	409.974	(1573.234, 3066.244)
		M_2	18585.459	548.410	133.428	(315.598, 837.713)
		M_3	18574.902	537.854	136.609	(331.680, 856.649)
		M_4	18040.594	3.545	54.883	(-97.074, 111.106)
	N=1500	M_0	27104.910	-	-	-
		M_1	30691.829	3586.919	471.146	(2746.449, 4566.627)
		M_2	27885.636	780.726	174.418	(478.045, 1135.086)
		M_3	27887.252	782.342	180.206	(379.828, 1142.793)
		M_4	27108.927	4.017	79.882	(-123.440, 144.897)
	N=2000	M_0	36066.012	-	-	-
		M_1	41063.885	4997.873	655.524	(3723.356, 6024.570)
		M_2	37172.099	1106.087	268.667	(570.621, 1602.051)
		M_3	37185.838	1119.826	268.117	(622.727, 1613.824)
		M_4	36054.359	-11.653	94.785	(-185.301, 153.191)
Prior II	N=1000	M_0	18037.049	-	-	-
		M_1	20810.895	2773.847	406.668	(2019.064, 3529.070)
		M_2	18585.459	548.410	133.428	(315.598, 837.713)
		M_3	18574.902	537.854	136.609	(331.680, 856.649)
		M_4	18040.594	3.545	54.883	(-97.074, 111.106)
	N=1500	M_0	29491.332	-	-	-
		M_1	31122.734	1631.401	450.888	(727.012, 2412.766)
		M_2	30256.928	765.596	161.661	(500.773, 1142.496)
		M_3	30271.349	780.016	166.609	(513.664, 1115.918)
		M_4	29489.946	-1.386	73.578	(-123.754, 137.684)
	N=2000	M_0	38544.984	-	-	-
		M_1	41501.122	2956.138	636.520	(1789.012, 4003.023)
		M_2	39638.304	1093.320	261.839	(507.750, 1551.527)
		M_3	39660.264	1115.280	258.207	(643.898, 1601.887)
		M_4	38527.284	-17.700	76.920	(-143.895, 147.313)

Table 4.2: Logarithm Bayes factor for simulation study

Prior	Logarithm BF	n=150	n=300	n=600
Type I	$\log B_{10}$	-1709.67	-2488.41	-3336.19
	$\log B_{20}$	-28.34	-38.56	-41.87
	$\log B_{30}$	-26.28	-36.33	-39.84
	$\log B_{40}$	-2.06	-2.02	-1.75
Type II	$\log B_{10}$	-1706.34	-2481.56	-3329.67
	$\log B_{20}$	-27.63	-37.87	-39.97
	$\log B_{30}$	-25.75	-35.35	-38.98
	$\log B_{40}$	-1.47	-1.63	-1.31

Table 4.3: The Bayesian estimates for within-groups model under accurate prior inputs

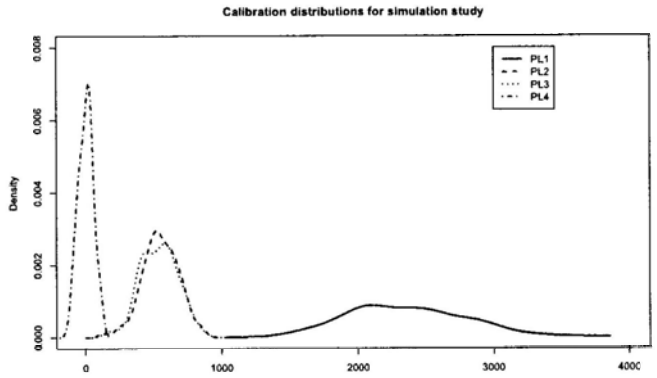
		$n = 1000$			$n = 1500$			$n = 2000$		
Para	TRUE	mean	RMS	BIAS	mean	RMS	BIAS	mean	RMS	BIAS
$\lambda_{1,21}$	0.800	0.804	0.032	0.004	0.798	0.022	0.002	0.802	0.025	0.002
$\lambda_{1,31}$	0.800	0.804	0.027	0.004	0.803	0.024	0.003	0.803	0.024	0.003
$\lambda_{1,52}$	0.800	0.803	0.037	0.003	0.805	0.036	0.005	0.801	0.025	0.001
$\lambda_{1,62}$	0.800	0.802	0.045	0.002	0.803	0.031	0.003	0.804	0.030	0.004
$\lambda_{1,83}$	0.800	0.801	0.038	0.001	0.808	0.035	0.008	0.803	0.031	0.003
$\lambda_{1,93}$	0.800	0.802	0.037	0.002	0.804	0.035	0.004	0.800	0.029	0.000
$\phi_{1,11}$	1.000	0.991	0.085	0.009	0.984	0.059	0.016	0.991	0.060	0.009
$\phi_{1,12}$	0.300	0.302	0.039	0.002	0.290	0.039	0.010	0.308	0.033	0.008
$\phi_{1,22}$	1.000	0.997	0.072	0.003	0.986	0.070	0.014	1.005	0.063	0.005
$\psi_{1\epsilon 1}$	0.640	0.640	0.042	0.000	0.639	0.042	0.001	0.642	0.037	0.002
$\psi_{1\epsilon 2}$	0.640	0.637	0.045	0.003	0.637	0.028	0.003	0.636	0.030	0.004
$\psi_{1\epsilon 3}$	0.640	0.637	0.042	0.003	0.632	0.027	0.008	0.641	0.028	0.001
$\psi_{1\epsilon 4}$	0.640	0.640	0.000	0.000	0.640	0.000	0.000	0.640	0.000	0.000
$\psi_{1\epsilon 5}$	0.640	0.640	0.035	0.000	0.638	0.030	0.002	0.636	0.027	0.004
$\psi_{1\epsilon 6}$	0.640	0.638	0.035	0.002	0.640	0.032	0.000	0.639	0.026	0.001
$\psi_{1\epsilon 7}$	0.640	0.640	0.000	0.000	0.640	0.000	0.000	0.640	0.000	0.000
$\psi_{1\epsilon 8}$	0.640	0.637	0.040	0.003	0.639	0.030	0.001	0.638	0.025	0.002
$\psi_{1\epsilon 9}$	0.640	0.637	0.042	0.003	0.641	0.031	0.001	0.643	0.029	0.003
$\psi_{1\delta}$	0.640	0.621	0.056	0.019	0.643	0.049	0.003	0.628	0.040	0.012
γ_{11}	0.600	0.604	0.051	0.004	0.602	0.041	0.002	0.605	0.035	0.005
γ_{12}	0.600	0.608	0.049	0.008	0.606	0.041	0.006	0.602	0.037	0.002
γ_{13}	-0.400	-0.400	0.042	0.000	-0.397	0.039	0.003	-0.398	0.040	0.002

Table 4.4: The Bayesian estimates for between-groups model under accurate prior inputs

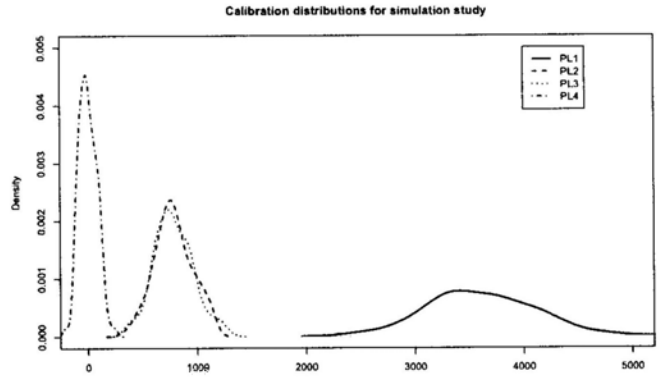
		$n = 1000$			$n = 1500$			$n = 2000$		
Para	TRUE	mean	RMS	BIAS	mean	RMS	BIAS	mean	RMS	BIAS
$\lambda_{2,21}$	0.800	0.813	0.100	0.013	0.829	0.087	0.029	0.812	0.065	0.012
$\lambda_{2,31}$	0.800	0.824	0.095	0.024	0.817	0.085	0.017	0.808	0.074	0.008
$\lambda_{2,52}$	0.800	0.816	0.081	0.016	0.807	0.073	0.007	0.799	0.053	0.001
$\lambda_{2,62}$	0.800	0.798	0.085	0.002	0.794	0.074	0.006	0.794	0.057	0.006
$\lambda_{2,83}$	0.800	0.798	0.077	0.002	0.804	0.059	0.004	0.797	0.054	0.003
$\lambda_{2,93}$	0.800	0.806	0.082	0.006	0.800	0.064	0.000	0.795	0.063	0.005
$\psi_{2\epsilon 1}$	0.360	0.366	0.047	0.006	0.356	0.047	0.004	0.355	0.038	0.005
$\psi_{2\epsilon 2}$	0.360	0.349	0.044	0.011	0.356	0.041	0.004	0.360	0.038	0.000
$\psi_{2\epsilon 3}$	0.360	0.343	0.050	0.017	0.343	0.044	0.017	0.351	0.038	0.009
$\psi_{2\epsilon 4}$	0.360	0.355	0.044	0.005	0.360	0.051	0.000	0.359	0.051	0.001
$\psi_{2\epsilon 5}$	0.360	0.359	0.043	0.001	0.358	0.042	0.002	0.359	0.036	0.001
$\psi_{2\epsilon 6}$	0.360	0.355	0.040	0.005	0.362	0.038	0.002	0.359	0.038	0.001
$\psi_{2\epsilon 7}$	0.360	0.351	0.045	0.009	0.353	0.044	0.007	0.358	0.043	0.002
$\psi_{2\epsilon 8}$	0.360	0.364	0.044	0.004	0.358	0.039	0.002	0.360	0.038	0.000
$\psi_{2\epsilon 9}$	0.360	0.357	0.042	0.003	0.360	0.050	0.000	0.358	0.043	0.002
$\psi_{2\delta}$	0.360	0.377	0.048	0.017	0.368	0.045	0.008	0.367	0.048	0.007
$\phi_{2,11}$	1.000	0.956	0.167	0.044	0.981	0.143	0.019	0.991	0.107	0.009
$\phi_{2,12}$	0.300	0.272	0.107	0.028	0.284	0.105	0.016	0.297	0.080	0.003
$\phi_{2,22}$	1.000	1.019	0.185	0.019	1.000	0.148	0.000	1.025	0.135	0.025
γ_{21}	0.600	0.547	0.122	0.053	0.564	0.100	0.036	0.597	0.078	0.003
γ_{22}	-0.600	-0.553	0.100	0.047	-0.568	0.088	0.032	-0.583	0.081	0.017
μ_1	0.000	-0.001	0.102	0.001	-0.006	0.077	0.006	0.005	0.075	0.005
μ_2	0.000	0.004	0.078	0.004	0.003	0.068	0.003	0.009	0.067	0.009
μ_3	0.000	-0.003	0.089	0.003	-0.008	0.074	0.008	0.002	0.061	0.002
μ_4	0.000	-0.008	0.093	0.008	0.007	0.085	0.007	0.004	0.070	0.004
μ_5	0.000	-0.013	0.085	0.013	0.005	0.066	0.005	-0.001	0.057	0.001
μ_6	0.000	-0.006	0.081	0.006	-0.001	0.077	0.001	-0.003	0.057	0.003
μ_7	0.000	0.005	0.107	0.005	-0.002	0.087	0.002	0.006	0.071	0.006
μ_8	0.000	-0.007	0.087	0.007	-0.002	0.075	0.002	0.008	0.057	0.008
μ_9	0.000	-0.005	0.091	0.005	0.003	0.075	0.003	-0.000	0.061	0.000

Table 4.5: Calibration summaries for real example

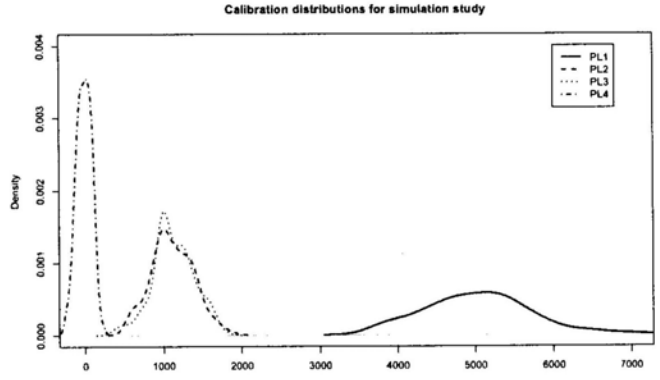
Model	$\mu(D_0.5)$	$SD(D_0.5)$	95% HPD
M_0	119.149	107.884	(-33.307, 358.124)
M_1	4412.300	320.986	(3857.586, 5108.685)
M_2	118.533	68.006	(-8.213, 228.110)
M_3	4350.957	318.336	(3820.036, 5005.845)
M_5	22.309	37.425	(-57.704, 88.144)



(a) N=1000

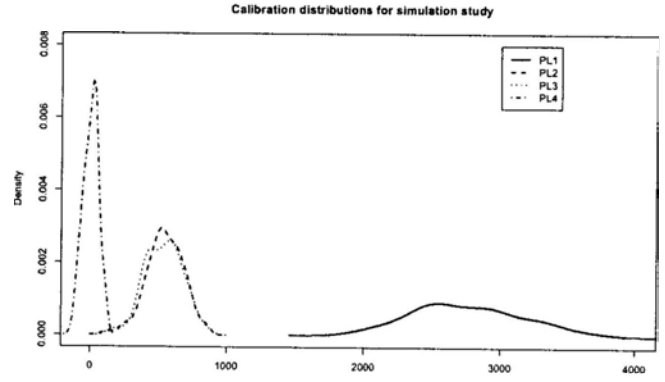


(b) N=1500

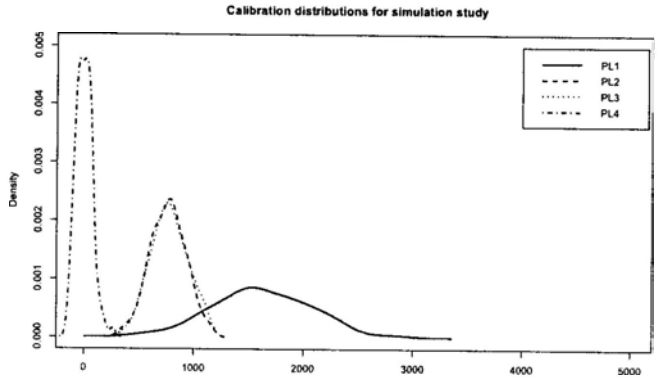


(c) N=2000

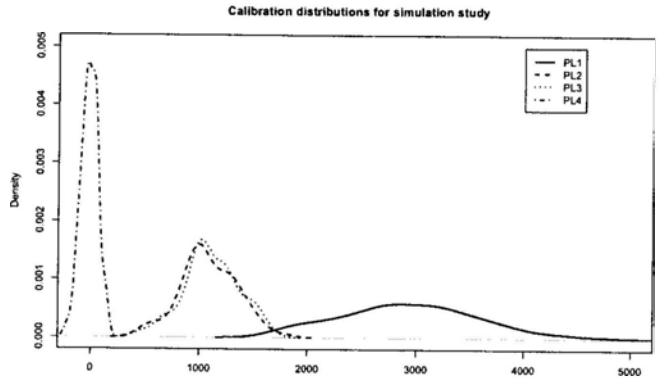
Figure 4.1: Calibration distributions under Prior I



(a) N=1000



(b) N=1500



(c) N=2000

Figure 4.2: Calibration distributions under Prior II

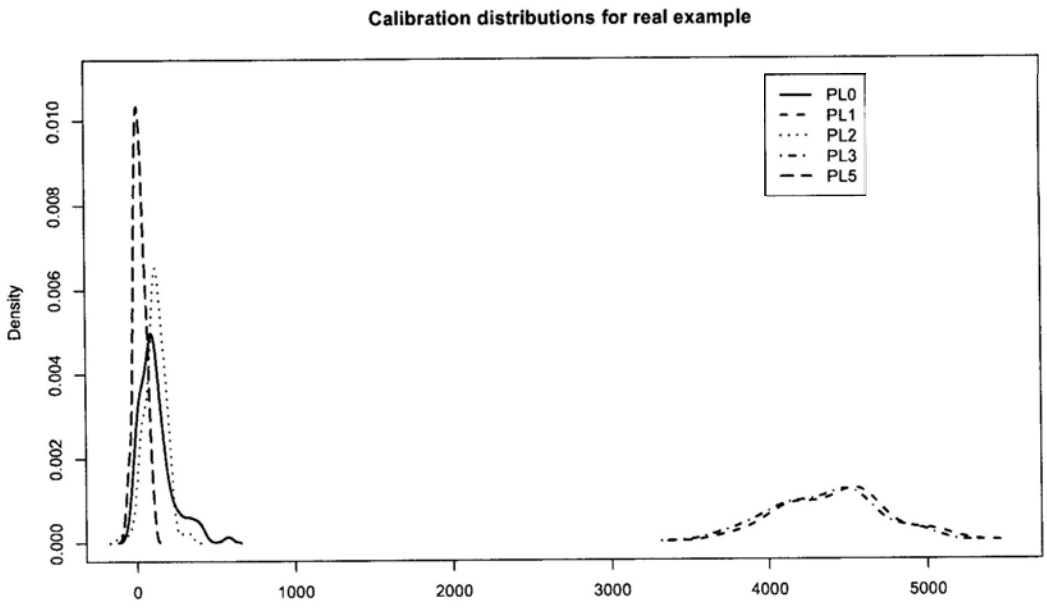


Figure 4.3: Example 2

Chapter 5

L_v Measure for Finite Mixture Structural Equation Models

5.1 Introduction

As mentioned in Chapter 4, heterogeneity of population is inevitable and is an important concern. In such cases, the observations may exhibit two possible kinds of heterogeneity. One kind of heterogeneous data, which has been discussed in Chapter 4, is drawn from a number of different groups (clusters) with a known hierarchical structure. The other kind is mixture data, in which the observations come from K populations with different distributions, and no information is available on which of the K populations that an individual observation belongs to. To analyze this kind of data, a finite mixture model (see Redner and Walker, 1984; Titterington et al., 1985) has been developed. In finite mixture models, K can be known or unknown. When K is known, the method of moments (Lindsay and Basak, 1993), the Bayesian approach (Diebolt and Robert, 1994; Robert, 1996; Lee, 2007), and the maximum likelihood (ML) method (Hathaway, 1985) can be used for the estimation of the model. When K is unknown, it becomes a difficult problem. To deal with this problem, Richardson and Green (1997) proposed a full Bayesian analysis on the basis of the reversible jump MCMC method, and treated the number of components K as ran-

dom; Lee and Song (2002, 2003a) and Song and Lee (2007, 2008) proposed a Bayesian approach for the analysis of finite mixture structural equation models (SEMs), and treated the problem of selecting the number of components as a model selection problem. Furthermore, they developed Bayes factor as a statistic for model selection. As we have discussed in previous chapters, Bayes factor has some advantages except for the difficulty in computation. Besides Bayes factor, deviance information criterion (DIC) is also a popular Bayesian method for model selection, and it can be obtained directly through WinBUGS for most of models. However, for mixture models, WinBUGs dose not produce the DIC values due to the reasons given by Celeux et al. (2006). Therefore, finding an efficient and simple method for model selection of mixture SEMs is important.

In this chapter, the L_v measure is proposed for model selection of finite mixture SEMs with unknown number of components. Similar as discussed in Lee and Song (2003a), we treat the problem of selecting the number of components as a model selection problem. Furthermore, the calibration of the L_v measure is also discussed. As the closed forms of the L_v measure and the calibration distribution for mixture SEMs cannot be obtained, MCMC methods are applied to the estimation of the L_v measure and the calibration distribution.

This chapter is organized as follows. In Section 5.2, a finite mixture of SEMs is defined. The L_v measure and its calibration distribution are discussed in Section 5.3. In Section 5.4, a simulation study is conducted to demonstrate the performance of the L_v measure for model selection of finite mixture SEMs. In Section 5.5, a real example is analyzed by using the L_v measure. A discussion is given in Section 5.6.

5.2 Model Description

Suppose $\boldsymbol{y}_i$ is a $p \times 1$ random vector corresponding to the ith observation in a random sample of size n, and the distribution of $\boldsymbol{y}_i$ is given by the following probability density function:

$$f(\boldsymbol{y}_i|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f_k(\boldsymbol{y}_i|\boldsymbol{u}_k, \boldsymbol{\theta}_k), i = 1, \cdots, n, \tag{5.1}$$

where K is a given integer, π_k is the unknown mixing proportion such that $\pi_k > 0$ and $\pi_1 + \cdots + \pi_K = 1.0$, $f_k(\boldsymbol{y}_i|\boldsymbol{u}_k, \boldsymbol{\theta}_k)$ is the multivariate normal density function with an unknown mean vector $\boldsymbol{u}_k$ and a general covariance structure $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_k)$ that depends on an unknown parameter vector $\boldsymbol{\theta}_k$, and $\boldsymbol{\theta}$ is the parameter vector that contains all unknown parameters in π_k, $\boldsymbol{u}_k$, and $\boldsymbol{\theta}_k$, $k = 1, \cdots, K$.

Similar as Lee (2007, chap. 11), the following well-known LISREL type model (Jöreskog and Sörbom, 1996) for the random vector $\boldsymbol{y}_i$ conditional on the kth component is considered. For the k-component, the measurement equation of the model is given by

$$\boldsymbol{y}_i = \boldsymbol{u}_k + \boldsymbol{\Lambda}_k \boldsymbol{\omega}_{ki} + \boldsymbol{\epsilon}_{ki}, \tag{5.2}$$

where $\boldsymbol{u}_k$ is the mean vector, $\boldsymbol{\Lambda}_k$ is the $p \times q$ factor loading matrix, $\boldsymbol{\omega}_{ki}$ is a random vector of latent variables, and $\boldsymbol{\epsilon}_{ki} \sim N[\boldsymbol{0}, \boldsymbol{\Psi}_k]$ is a random vector of residuals, where $\boldsymbol{\Psi}_k$ is a diagonal matrix. It is assumed that $\boldsymbol{\omega}_{ki}$ and $\boldsymbol{\epsilon}_{ki}$ are independent. Moreover, let $\boldsymbol{\omega}_{ki} = (\boldsymbol{\eta}_{ki}^T, \boldsymbol{\xi}_{ki}^T)^T$. The structural equation of the model, which describes the relationships among latent variables, is defined as

$$\boldsymbol{\eta}_{ki} = \boldsymbol{\Pi}_k \boldsymbol{\eta}_{ki} + \boldsymbol{\Gamma}_k \boldsymbol{\xi}_{ki} + \boldsymbol{\delta}_{ki}, \tag{5.3}$$

where $\boldsymbol{\eta}_{ki}$ and $\boldsymbol{\xi}_{ki}$ are $q_1 \times 1$ and $q_2 \times 1$ subvectors of $\boldsymbol{\omega}_{ki}$ respectively; $\boldsymbol{\Pi}_k$ and $\boldsymbol{\Gamma}_k$ are unknown parameter matrices such that $\boldsymbol{\Pi}_{0k}^{-1} = (\boldsymbol{I} - \boldsymbol{\Pi}_k)^{-1}$ exists, and $|\boldsymbol{\Pi}_{0k}|$ is independent of the elements in $\boldsymbol{\Pi}_k$. $\boldsymbol{\delta}_{ki}$ is a random vector of residuals that is independent of $\boldsymbol{\xi}_{ki}$; Let the distributions of $\boldsymbol{\xi}_k$ and $\boldsymbol{\delta}_k$

be $N[\mathbf{0}, \boldsymbol{\Phi}_k]$ and $N[\mathbf{0}, \boldsymbol{\Psi}_{\delta k}]$, respectively, where $\boldsymbol{\Psi}_{\delta k}$ is a diagonal matrix. The parameter vector $\boldsymbol{\theta}_k$ in the k-component contains all the unknown parameters in $\boldsymbol{\Lambda}_k, \boldsymbol{\Pi}_k, \boldsymbol{\Gamma}_k, \boldsymbol{\Phi}_k, \boldsymbol{\Psi}_{\delta k}$, and $\boldsymbol{\Psi}_k$. The covariance structure of $\boldsymbol{\omega}_{ki}$ is given by

$$\boldsymbol{\Sigma}_{\omega k} = \begin{bmatrix} \boldsymbol{\Pi}_{0k}^{-1}(\boldsymbol{\Gamma}_k \boldsymbol{\Phi}_k \boldsymbol{\Gamma}_k^T + \boldsymbol{\Psi}_{\delta k})(\boldsymbol{\Pi}_{0k}^{-1})^T & \boldsymbol{\Pi}_{0k}^{-1}\boldsymbol{\Gamma}_k \boldsymbol{\Phi}_k \\ \boldsymbol{\Phi}_k \boldsymbol{\Gamma}_k^T (\boldsymbol{\Pi}_{0k}^{-1})^T & \boldsymbol{\Phi}_k \end{bmatrix} \tag{5.4}$$

The covariance structure for $\boldsymbol{y}_i$ under the kth component is $\boldsymbol{\Sigma}_k(\boldsymbol{\theta}_k) = \boldsymbol{\Lambda}_k \boldsymbol{\Sigma}_{\omega k} \boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi}_k$.

As the mixture model defined in (5.2) is invariant with respect to permutation of labels $k = 1, \cdots, K$, adoption of an unique labeling for identifiability is important. The method proposed in Roeder and Wasserman (1997) and Zhu and Lee (2001) is proposed to impose the ordering $u_{1,1} < \cdots < u_{K,1}$ for solving the label switching problem (jumping between the various labeling subspace), where $u_{k,1}$ is the first element of the mean vector $\boldsymbol{u}_k$. This method works well if $u_{1,1} < \cdots < u_{K,1}$ are well separated. However, if $u_{1,1} < \cdots < u_{K,1}$ are close to each other, it may not be able to eliminate the label switching, and may give bias results. Hence, it is important to find an appropriate identifiability constraint. Here, the random permutation sampler that is developed by Frühwirth-Schnatter (2001) is applied to finding the suitable identifiability constraints. Moreover, for each $k = 1, \cdots, K$, the structural parameters in the covariance matrix $\boldsymbol{\Sigma}_k$ corresponding to the model defined by equations (5.2) and (5.3) are not identified. This problem is solved by the common method in structural equation modeling by fixing appropriate elements in $\boldsymbol{\Lambda}_k, \boldsymbol{\Pi}_k$, and/or $\boldsymbol{\Gamma}_k$ at preassigned values that are chosen on problem-by-problem basis. For clear presentation of the Bayesian method, we assume that all the unknown parameters in the model are identified. Furthermore, inspired by many works in finite mixture models, a group label w_i for the ith observation $\boldsymbol{y}_i$ is considered as a latent allocation variable, and is assumed independently drawn from the

following distribution:

$$p(\mathrm{w}_i = k) = \pi_k, \text{ for } k = 1, \cdots, K. \tag{5.5}$$

We treat w_i as a categorical variable, and to apply L_v measure to the proposed mixture SEMs, a new vector w_i^* is introduced. Specifically, $\mathsf{w}_i^* = (\mathrm{w}_{i1}^*, \cdots, \mathrm{w}_{iK}^*)^T$, and for $k = 1, \cdots, K$,

$$\mathrm{w}_{ik}^* = \begin{cases} 1, & \text{if } \mathrm{w}_i = k, \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

We denote the proposed model by M. For convenience, let $\boldsymbol{Y}^{obs} = (\boldsymbol{y}_1^{obs}, \cdots, \boldsymbol{y}_n^{obs})$ be the matrix of observed data, where $\boldsymbol{y}_i^{obs} = (y_{i1}, \cdots, y_{ip})^T (i = 1, \cdots, n)$ is the ith column of $\boldsymbol{Y}^{obs}$, and $\boldsymbol{Y}^{rep} = (\boldsymbol{y}_1^{rep}, \cdots, \boldsymbol{y}_n^{rep})$, which has the same distribution with $\boldsymbol{Y}^{obs}$, be the replicated data, where $\boldsymbol{y}_i^{rep} = (y_{i1}^{rep}, \cdots, y_{ip}^{rep})^T$. Let $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ be the matrix of latent variables, $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ be the submatrices of $\boldsymbol{\Omega}$ corresponding to $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, respectively. Let $\boldsymbol{W} = (\mathrm{w}_1, \cdots, \mathrm{w}_n)$ be the observation of the latent allocation variable, where $\mathrm{w}_i \in \{1, \cdots, K\}$. Furthermore, let $\boldsymbol{W}^* = (\mathsf{w}_1^*, \cdots, \mathsf{w}_n^*)$, and $\boldsymbol{W}^{*rep} = (\mathsf{w}_1^{*rep}, \cdots, \mathsf{w}_n^{*rep})$ be the replication which has the same distribution with $\boldsymbol{W}^*$.

5.3 L_v Measure for Finite Mixture SEMs

5.3.1 Definition of the L_v Measure

For the proposed mixture model, there are two important issues. One is prediction, and the other is classification. Therefore, two measures are considered for the proposed mixture SEMs. First, we propose a measure that considers the accuracy of the prediction. The L_v measure for the proposed model is defined as follows:

$$L_{1v}(\boldsymbol{Y}^{obs}, M) = \sum_{i=1}^{n} \left[\mathrm{tr}(Var(\boldsymbol{y}_i^{rep} | \boldsymbol{Y}^{obs}, M)) + v(\boldsymbol{y}_i^{obs} - \boldsymbol{\mu}_i)^T (\boldsymbol{y}_i^{obs} - \boldsymbol{\mu}_i) \right], \tag{5.7}$$

where $\boldsymbol{\mu}_i = E(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M)$, in which the conditional expectation is taken with respect to the predicted posterior distribution $p(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M)$ defined by

$$\begin{aligned} & p(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) \\ & = \int \sum_{k=1}^{K} p(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i = k, M) p(\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i = k|\boldsymbol{Y}^{obs}, M) d\boldsymbol{\theta} d\boldsymbol{\xi}_i. \end{aligned} \tag{5.8}$$

Therefore, the conditional expectation $\boldsymbol{\mu}_i$ is given as follows:

$$\begin{aligned} \boldsymbol{\mu}_i &= E(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) = \int \boldsymbol{y}_i^{rep} p(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) d\boldsymbol{y}_i^{rep} \\ &= \int \sum_{k=1}^{K} \int \boldsymbol{y}_i^{rep} p(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i = k, M) d\boldsymbol{y}_i^{rep} p(\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i = k|\boldsymbol{Y}^{obs}, M) d\boldsymbol{\theta} d\boldsymbol{\xi}_i \\ &= E_{\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i|\boldsymbol{Y}^{obs}, M} \left[E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i, M) \right], \end{aligned} \tag{5.9}$$

and

$$E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i = k, M) = \boldsymbol{u}_k + \boldsymbol{\Lambda}_{k\eta}(\boldsymbol{\Pi}_{0k}^{-1}\boldsymbol{\Gamma}_k\boldsymbol{\xi}_{ki}) + \boldsymbol{\Lambda}_{k\xi}\boldsymbol{\xi}_{ki},$$

where $\boldsymbol{\Lambda}_{k\eta}$ and $\boldsymbol{\Lambda}_{k\xi}$ are submatrices of $\boldsymbol{\Lambda}_k$ corresponding to $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, respectively. For the conditional variance of $\boldsymbol{y}_i^{rep}$, we have

$$\begin{aligned} & Var(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) \\ & = E\{(\boldsymbol{y}_i^{rep})(\boldsymbol{y}_i^{rep})^T|\boldsymbol{Y}^{obs}, M\} - E(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) E(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M)^T, \end{aligned}$$

similarly, we can get

$$\begin{aligned} & E\{(\boldsymbol{y}_i^{rep})(\boldsymbol{y}_i^{rep})^T|\boldsymbol{Y}^{obs}, M\} = \int (\boldsymbol{y}_i^{rep})(\boldsymbol{y}_i^{rep})^T p(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) d\boldsymbol{y}_i^{rep} \\ & = E_{\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i|\boldsymbol{Y}^{obs}, M} \left[E\{(\boldsymbol{y}_i^{rep})(\boldsymbol{y}_i^{rep})^T|\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i, M\} \right] \\ & = E_{\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i|\boldsymbol{Y}^{obs}, M} \left[Var(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i, M) \right] \\ & \quad + E_{\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i|\boldsymbol{Y}^{obs}, M} \left[E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i, M) E(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i, M)^T \right], \end{aligned} \tag{5.10}$$

and

$$Var(\boldsymbol{y}_i^{rep}|\boldsymbol{\theta}, \boldsymbol{\xi}_i, \mathrm{w}_i = k, M) = (\boldsymbol{\Lambda}_{k\eta}\boldsymbol{\Pi}_{0k}^{-1})\boldsymbol{\Psi}_{\delta k}(\boldsymbol{\Lambda}_{k\eta}\boldsymbol{\Pi}_{0k}^{-1})^T + \boldsymbol{\Psi}_k. \tag{5.11}$$

Now, we consider the accuracy of classification. As the higher the accuracy of classification, the lower the rate of misclassification. Therefore, instead of considering the accuracy of the classification, we consider the loss of misclassification. Again, the L_v measure is applied, and it can be defined by

$$L_{2v}(\boldsymbol{Y}^{obs}, M) = \sum_{i=1}^{n} \left[\text{tr}(Var(\mathsf{w}_i^{*rep}|\boldsymbol{Y}^{obs}, M)) + v(\mathsf{w}_i^* - \boldsymbol{\mu}_i^*)(\mathsf{w}_i^* - \boldsymbol{\mu}_i^*)^T\right], \tag{5.12}$$

where $\boldsymbol{\mu}_i^* = E(\mathsf{w}_i^{*rep}|\boldsymbol{Y}^{obs}, M) = (E(\mathrm{w}_{i1}^{*rep}|\boldsymbol{Y}^{obs}, M), \cdots, E(\mathrm{w}_{iK}^{*rep}|\boldsymbol{Y}^{obs}, M))^T$, and for $k = 1, \cdots, K$,

$$\begin{aligned} E(\mathrm{w}_{ik}^{*rep}|\boldsymbol{Y}^{obs}, M) &= p(\mathrm{w}_{ik}^{*rep} = 1|\boldsymbol{Y}^{obs}, M) = p(\mathrm{w}_i^{rep} = k|\boldsymbol{Y}^{obs}, M) \\ &= \int p(\mathrm{w}_i^{rep} = k|\boldsymbol{u}_k, \boldsymbol{\theta}_k, \boldsymbol{Y}^{obs}, M) p(\boldsymbol{u}_k, \boldsymbol{\theta}_k|\boldsymbol{Y}^{obs}, M) d\boldsymbol{u}_k d\boldsymbol{\theta}_k, \end{aligned} \tag{5.13}$$

where $p(\mathrm{w}_i^{rep} = k|\boldsymbol{u}_k, \boldsymbol{\theta}_k, \boldsymbol{Y}^{obs}, M) = \frac{\pi_k f_k(\boldsymbol{y}_i^{obs}|\boldsymbol{u}_k, \boldsymbol{\theta}_k)}{f(\boldsymbol{y}_i^{obs}|\boldsymbol{\theta})}$. After obtaining the conditional expectation, the kth diagonal element of the conditional covariance matrix $Var(\mathsf{w}_i^{*rep}|\boldsymbol{Y}^{obs}, M)$ can be obtained easily as: $Var(\mathsf{w}_{ik}^{*rep}|\boldsymbol{Y}^{obs}, M) = p(\mathrm{w}_{ik}^{*rep} = 1|\boldsymbol{Y}^{obs}, M)(1 - p(\mathrm{w}_{ik}^{*rep} = 1|\boldsymbol{Y}^{obs}, M))$. To calculate the measure defined by equation (5.12), w_i^* is replaced by the Bayesian estimates, we will discuss it later.

Combine equations (5.7) and (5.12), L_v measure for the proposed mixture model is given by

$$\begin{aligned} L_v(\boldsymbol{Y}^{obs}, M) = &\sum_{i=1}^{n} \left[\text{tr}(Var(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M)) + v(\boldsymbol{y}_i^{obs} - \boldsymbol{\mu}_i)^T(\boldsymbol{y}_i^{obs} - \boldsymbol{\mu}_i)\right] \\ &+ \sum_{i=1}^{n} \left[\text{tr}(Var(\mathsf{w}_i^{*rep}|\boldsymbol{Y}^{obs}, M)) + v(\mathsf{w}_i^* - \boldsymbol{\mu}_i^*)^T(\mathsf{w}_i^* - \boldsymbol{\mu}_i^*)\right]. \end{aligned} \tag{5.14}$$

From the above discussion, the most important issue in calculating the L_v measure is to simulate observations from the joint posterior distribution

$p(\boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{W}|\boldsymbol{Y}^{obs}, M)$. According to the definition of the model, it is impossible to get a closed form of this posterior distribution. Therefore, MCMC methods are used.

5.3.2 Computation of the L_v Measure

Suppose $\{\boldsymbol{\theta}^{(r)}, \boldsymbol{W}^{(r)}, \boldsymbol{\Omega}^{(r)}; r = 1, \cdots, R\}$ is an MCMC sequence generated from $p(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{\Omega}|\boldsymbol{Y}^{obs}, M)$. Based on the posterior observations, the Bayesian estimation of the unknown parameters, latent variables, and the allocation variables of the proposed mixture model can be obtained, and then the L_v measure can be easily computed. Specifically, for the proposed finite mixture SEMs, L_v measure can be calculated as:

$$\begin{aligned} L_v(\boldsymbol{Y}^{obs}, M) = & \sum_{i=1}^{n} \left[\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_i) + v(\boldsymbol{y}_i^{obs} - \widehat{\boldsymbol{\mu}}_i)^T (\boldsymbol{y}_i^{obs} - \widehat{\boldsymbol{\mu}}_i) \right] \\ & + \sum_{i=1}^{n} \left[(\boldsymbol{p}_i^*)^T (\mathbf{1} - \boldsymbol{p}_i^*) + v(\boldsymbol{p}_i^* - \boldsymbol{z}_i^*)^T (\boldsymbol{p}_i^* - \boldsymbol{z}_i^*) \right]. \end{aligned} \tag{5.15}$$

In the first summation,

$$\begin{aligned} \widehat{\boldsymbol{\mu}}_i &= \frac{1}{R} \sum_{r=1}^{R} \boldsymbol{\mu}_{k^{(r)}i}^{(r)}, \\ \widehat{\boldsymbol{\Sigma}}_i &= \widehat{Var}(\boldsymbol{y}_i^{rep}|\boldsymbol{Y}^{obs}, M) \\ &= \frac{1}{R} \sum_{r=1}^{R} \Big\{ \boldsymbol{\Lambda}_{k^{(r)}\eta}^{(r)} (\boldsymbol{\Pi}_{0k^{(r)}}^{(r)})^{-1} \boldsymbol{\Psi}_{\delta k^{(r)}}^{(r)} (\boldsymbol{\Lambda}_{k^{(r)}\eta}^{(r)} (\boldsymbol{\Pi}_{0k^{(r)}}^{(r)})^{-1})^T + \boldsymbol{\Psi}_{k^{(r)}}^{(r)} \\ &\quad + (\boldsymbol{\mu}_{k^{(r)}i}^{(r)})^T (\boldsymbol{\mu}_{k^{(r)}i}^{(r)}) \Big\} - \widehat{\boldsymbol{\mu}}_i \widehat{\boldsymbol{\mu}}_i^T, \end{aligned}$$

where $k^{(r)} = \mathrm{w}_i^{(r)}$, and

$$\boldsymbol{\mu}_{k^{(r)}i}^{(r)} = \boldsymbol{u}_{k^{(r)}}^{(r)} + \boldsymbol{\Lambda}_{k^{(r)}\eta}^{(r)} (\boldsymbol{\Pi}_{0k^{(r)}}^{(r)})^{-1} \boldsymbol{\Gamma}_{k^{(r)}}^{(r)} \boldsymbol{\xi}_{k^{(r)}i}^{(r)} + \boldsymbol{\Lambda}_{k^{(r)}\xi}^{(r)} \boldsymbol{\xi}_{k^{(r)}i}^{(r)}.$$

In the second summation, $\mathbf{1}$ is a K-dimensional vector with all elements being 1, $\boldsymbol{p}_i^* = (p_{i1}^*, \cdots, p_{iK}^*)^T$ and $\boldsymbol{z}_i^* = (z_{i1}^*, \cdots, z_{iK}^*)$, where for $k = 1, \cdots, K$,

$$p_{ik}^* = \frac{1}{R} \sum_{r=1}^{R} I(\mathrm{w}_i^{(r)} = k), \text{ and } z_{ik}^* = \begin{cases} 1, & \text{if } \max(p_i^*) = p_{ik}^*. \\ 0, & \text{otherwise.} \end{cases}$$

Gibbs sampler algorithm (see Geman and Geman, 1984) is applied to generate observations required in computing the L_v measure. With current values $\boldsymbol{\theta}^{(r)}, \boldsymbol{\Omega}^{(r)}$, and $\boldsymbol{W}^{(r)}$, the Gibbs sampler is implemented as follows:

Step (a) Generate $\boldsymbol{W}^{(r+1)}$ from $p(\boldsymbol{W}|\boldsymbol{\theta}^{(r)}, \boldsymbol{Y}^{obs}, M)$.

Step (b) Generate $\boldsymbol{\Omega}^{(r+1)}$ from $p(\boldsymbol{\Omega}|\boldsymbol{\theta}^{(r)}, \boldsymbol{W}^{(r+1)}, \boldsymbol{Y}^{obs}, M)$.

Step (c) Generate $\boldsymbol{\theta}^{(r+1)}$ from $p(\boldsymbol{\theta}|\boldsymbol{\Omega}^{(r+1)}, \boldsymbol{W}^{(r+1)}, \boldsymbol{Y}^{obs}, M)$.

5.3.3 Full Conditional Distributions

In this part, the full conditional distributions required in the Gibbs sampler are discussed. First, we consider the conditional distribution associated with Step (a). As w_i's are mutually independent,

$$p(\boldsymbol{W}|\boldsymbol{Y}^{obs}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(\mathrm{w}_i|\boldsymbol{y}_i^{obs}, \boldsymbol{\theta}). \tag{5.16}$$

and

$$\begin{aligned} p(\mathrm{w}_i = k|\boldsymbol{y}_i^{obs}, \boldsymbol{\theta}) &= \frac{p(\mathrm{w}_i = k, \boldsymbol{y}_i^{obs}|\boldsymbol{\theta})}{p(\boldsymbol{y}_i^{obs}|\boldsymbol{\theta})} \\ &= \frac{p(\mathrm{w}_i = k|\boldsymbol{\pi})p(\boldsymbol{y}_i^{obs}|\mathrm{w}_i = k, \boldsymbol{\theta})}{p(\boldsymbol{y}_i^{obs}|\boldsymbol{\theta})} \\ &= \frac{\boldsymbol{\pi}_k f_k(\boldsymbol{y}_i^{obs}|\boldsymbol{u}_k, \boldsymbol{\theta}_k)}{p(\boldsymbol{y}_i^{obs}|\boldsymbol{\theta})}, \end{aligned} \tag{5.17}$$

where $f_k(\boldsymbol{y}_i^{obs}|\boldsymbol{u}_k, \boldsymbol{\theta}_k)$ is a multivariate normal density function with mean vector $\boldsymbol{u}_k$ and covariance matrix $\boldsymbol{\Sigma}_k(\boldsymbol{\theta}_k)$. Therefore, the conditional distribution of $\boldsymbol{W}$ given $\boldsymbol{Y}^{obs}$ and $\boldsymbol{\theta}$ can be derived from (5.17). Drawing observations from this conditional distribution is straightforward.

Now, we consider the conditional distribution $p(\boldsymbol{\Omega}|\boldsymbol{Y}^{obs}, M, \boldsymbol{W}, \boldsymbol{\theta})$ required in Step (b). As $\boldsymbol{\omega}_i$ are mutually independent given w_i and $\boldsymbol{\theta}$, we

have

$$
\begin{aligned}
& p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \mathrm{w}_i = k, \boldsymbol{Y}^{obs}) \\
&= \prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\boldsymbol{\theta}, \mathrm{w}_i = k, \boldsymbol{y}_i^{obs}) \\
&\propto \prod_{i=1}^{n} p(\boldsymbol{y}_i^{obs}|\boldsymbol{\theta}, \boldsymbol{\omega}_i, \mathrm{w}_i = k) p(\boldsymbol{\omega}_i|\boldsymbol{\theta}, \mathrm{w}_i = k), \qquad (5.18)
\end{aligned}
$$

where $p(\boldsymbol{y}_i^{obs}|\boldsymbol{\theta}, \boldsymbol{\omega}_i, \mathrm{w}_i = k) \triangleq N[\boldsymbol{u}_k + \boldsymbol{\Lambda}_k\boldsymbol{\omega}_i, \boldsymbol{\Psi}_k]$ and $p(\boldsymbol{\omega}_i|\boldsymbol{\theta}, \mathrm{w}_i = k) \triangleq N[\boldsymbol{0}, \boldsymbol{\Sigma}_{\omega k}]$. Let $\boldsymbol{C}_k = \boldsymbol{\Sigma}_{\omega k}^{-1} + \boldsymbol{\Lambda}_k^T\boldsymbol{\Psi}_k^{-1}\boldsymbol{\Lambda}_k$, it can be shown that

$$
p(\boldsymbol{\omega}_i|\boldsymbol{\theta}, \mathrm{w}_i = k, \boldsymbol{y}_i^{obs}) \triangleq N[\boldsymbol{C}_k^{-1}\boldsymbol{\Lambda}_k^T\boldsymbol{\Psi}_k^{-1}(\boldsymbol{y}_i^{obs} - \boldsymbol{\mu}_k), \boldsymbol{C}_k^{-1}]. \qquad (5.19)
$$

Then the conditional distribution of $p(\boldsymbol{\Omega}|\boldsymbol{Y}^{obs}, \boldsymbol{\theta}, \boldsymbol{W})$ can be obtained.

Finally, we consider the conditional distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}, \boldsymbol{W})$. As discussed in Lee (2007, chap. 11), the conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{Y}^{obs}, \boldsymbol{\Omega}$, and $\boldsymbol{W}$ is complicated, but the complexity can be reduced by assuming some mild conditions on the prior distribution of $\boldsymbol{\theta}$. Let $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_K)$, $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_K)$, $\boldsymbol{\theta}_y$ be the vector that contains all the unknown parameters in $\{\boldsymbol{\Lambda}_k, \boldsymbol{\Psi}_k; k = 1, \cdots, K\}$ relating to the measurement equation, and $\boldsymbol{\theta}_\omega$ be the vector that contains all the unknown parameters in $\{\boldsymbol{\Phi}_k, \boldsymbol{\Lambda}_{\omega k}, \boldsymbol{\Psi}_{\delta k}; k = 1, \cdots, K\}$ relating the structural equation. The assumptions are given as follows:

(1) Assume that the prior distributions of the mixing proportion $\boldsymbol{\pi}$ is independent of the prior distributions of $\boldsymbol{u}, \boldsymbol{\theta}_y$, and $\boldsymbol{\theta}_\omega$.

(2) Assume that the prior distribution of the mean vector $\boldsymbol{u}$ can be taken to be independent of the prior distributions of the parameters $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_\omega$ in the covariance structures.

(3) Assume that the prior distributions of $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_\omega$ are independent.

From the above assumptions, we get

$$
p(\boldsymbol{\theta}) = p(\boldsymbol{\pi}, \boldsymbol{u}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_\omega) = p(\boldsymbol{\pi})p(\boldsymbol{u})p(\boldsymbol{\theta}_y)p(\boldsymbol{\theta}_\omega).
$$

Furthermore, from the definition of the model and the properties of $\boldsymbol{\theta}$, $\boldsymbol{\Omega}$,

and $\boldsymbol{W}$, we have

$$
\begin{aligned}
&p(\boldsymbol{W}|\boldsymbol{\theta}) = p(\boldsymbol{W}|\boldsymbol{\pi}), \\
&p(\boldsymbol{\Omega}, \boldsymbol{Y}^{obs}|\boldsymbol{W}, \boldsymbol{\theta}) = p(\boldsymbol{Y}^{obs}|\boldsymbol{u}, \boldsymbol{\theta}_y, \boldsymbol{\Omega}, \boldsymbol{W}) p(\boldsymbol{\Omega}|\boldsymbol{\theta}_\omega, \boldsymbol{W}), \\
&p(\boldsymbol{\Omega}|\boldsymbol{W}, \boldsymbol{\theta}_\omega) = p(\boldsymbol{\Omega}|\boldsymbol{\theta}_\omega).
\end{aligned}
$$

Therefore, the joint conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{W}, \boldsymbol{\Omega}$, and $\boldsymbol{Y}^{obs}$ can be expressed as

$$
\begin{aligned}
&p(\boldsymbol{\theta}|\boldsymbol{W}, \boldsymbol{\Omega}, \boldsymbol{Y}^{obs}) = p(\boldsymbol{\pi}, \boldsymbol{u}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_\omega|\boldsymbol{\Omega}, \boldsymbol{W}, \boldsymbol{Y}^{obs}) \\
&\propto p(\boldsymbol{\pi}) p(\boldsymbol{u}) p(\boldsymbol{\theta}_y) p(\boldsymbol{\theta}_\omega) p(\boldsymbol{W}, \boldsymbol{\Omega}, \boldsymbol{Y}^{obs}|\boldsymbol{\theta}) \\
&\propto p(\boldsymbol{\pi}) p(\boldsymbol{u}) p(\boldsymbol{\theta}_y) p(\boldsymbol{\theta}_\omega) p(\boldsymbol{W}|\boldsymbol{\pi}) p(\boldsymbol{\Omega}|\boldsymbol{\theta}_\omega) p(\boldsymbol{Y}^{obs}|\boldsymbol{u}, \boldsymbol{\theta}_y, \boldsymbol{\Omega}, \boldsymbol{W}) \\
&= \left[p(\boldsymbol{\pi}) p(\boldsymbol{W}|\boldsymbol{\pi})\right] \left[p(\boldsymbol{u}) p(\boldsymbol{\theta}_y) p(\boldsymbol{Y}^{obs}|\boldsymbol{u}, \boldsymbol{\theta}_y, \boldsymbol{\Omega}, \boldsymbol{W})\right] \left[p(\boldsymbol{\theta}_\omega) p(\boldsymbol{\Omega}|\boldsymbol{\theta}_\omega\right].
\end{aligned}
\tag{5.20}
$$

Thus we can consider the marginal densities $p(\boldsymbol{\pi}|\cdot)$, $p(\boldsymbol{u}, \boldsymbol{\theta}_y|\cdot)$, and $p(\boldsymbol{\theta}_\omega|\cdot)$ separately.

First, we consider the posterior distribution $p(\boldsymbol{\pi}|\cdot)$. Like many Bayesian analyses in SEMs, we take the symmetric Dirichlet distribution as the prior distribution of $\boldsymbol{\pi}$. Specifically, $\boldsymbol{\pi} \sim D(\alpha, \cdots, \alpha)$ with probability density function given by

$$
p(\boldsymbol{\pi}) = \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^k} \pi_1^\alpha \cdots \pi_K^\alpha,
$$

where $\Gamma(\cdot)$ is the Gamma function. From equation (5.5), $p(\boldsymbol{W}|\boldsymbol{\pi}) \propto \prod_{k=1}^{K} \pi_k^{n_k}$, then the conditional distribution for $\boldsymbol{\pi}$ can be given by

$$
p(\boldsymbol{\pi}|\cdot) \propto p(\boldsymbol{\pi}) p(\boldsymbol{W}|\boldsymbol{\pi}) \propto \prod_{k=1}^{K} \pi_k^{n_k+\alpha}, \tag{5.21}
$$

where n_k is the total number of i such that $\mathrm{w}_i = k$. Thus, $p(\boldsymbol{\pi}|\cdot) \triangleq D(\alpha + n_1, \cdots, \alpha + n_K)$.

Now, we consider the posterior distributions of the parameters in $\boldsymbol{u}$, $\boldsymbol{\theta}_y$, and $\boldsymbol{\theta}_\omega$. We use the same conjugate prior distributions as given in other Bayesian literature. Let $\boldsymbol{Y}_k^{obs}$ and $\boldsymbol{\Omega}_k$ be the submatrices of $\boldsymbol{Y}^{obs}$ and $\boldsymbol{\Omega}$ such that all the ith columns with $\mathrm{w}_i \neq k$ are deleted. Moreover, let $\boldsymbol{\Omega}_{1k}$

and $\boldsymbol{\Omega}_{2k}$ be the submatrices of $\boldsymbol{\Omega}_k$ related to $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$, respectively. Then it is natural to assume that for $k \neq h$, $(\boldsymbol{u}_k, \boldsymbol{\theta}_{yk}, \boldsymbol{\theta}_{\omega k})$ and $(\boldsymbol{u}_h, \boldsymbol{\theta}_{yh}, \boldsymbol{\theta}_{\omega h})$ are independent. Thus given $\boldsymbol{W}$, we have

$$\begin{aligned} & p(\boldsymbol{u}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_\omega | \boldsymbol{\Omega}, \boldsymbol{W}, \boldsymbol{Y}^{obs}) \\ & \propto \prod_{k=1}^{K} p(\boldsymbol{u}_k) p(\boldsymbol{\theta}_{yk}) p(\boldsymbol{\theta}_{\omega k}) p(\boldsymbol{Y}_k^{obs} | \boldsymbol{u}_k, \boldsymbol{\theta}_{yk}, \boldsymbol{\Omega}_k) p(\boldsymbol{\Omega}_k | \boldsymbol{\theta}_{\omega k}). \end{aligned} \tag{5.22}$$

So, we can treat the product in (5.22) separately with each k. To get the posterior distributions of the parameters involved in $\boldsymbol{u}_k$, $\boldsymbol{\theta}_{yk}$ and $\boldsymbol{\theta}_{\omega k}$ for $k = 1, \cdots, K$, proper prior distributions are required. In this chapter, conjugate priors are used. Specifically, we consider the following conjugate prior distribution for $\boldsymbol{u}_k$:

$$p(\boldsymbol{u}_k) \triangleq N[\boldsymbol{u}_0, \boldsymbol{\Sigma}_0].$$

For the parameters involved in $\boldsymbol{\theta}_{yk}$, the following conjugate priors are considered: for $j = 1, \cdots, p$,

$$p(\boldsymbol{\Lambda}_{kj} | \psi_{kj}) \triangleq N[\boldsymbol{\Lambda}_{0kj}, \psi_{kj} \boldsymbol{H}_{0ykj}], \; p(\psi_{kj}^{-1}) \triangleq Gamma(\alpha_{0\epsilon k}, \beta_{0\epsilon k}),$$

where $\boldsymbol{\Lambda}_{kj}$ is the jth row of $\boldsymbol{\Lambda}_k$, ψ_{kj} is the jth diagonal element of $\boldsymbol{\Psi}_k$; $\boldsymbol{u}_0, \boldsymbol{\Lambda}_{0kj}, \alpha_{0\epsilon k}, \beta_{0\epsilon k}$, and the positive definite matrices $\boldsymbol{\Sigma}_0$ and $\boldsymbol{H}_{0ykj}$ are all given hyperparameters. By some derivations, the posterior distributions for them can be given by

$$\begin{aligned} & [\boldsymbol{u}_k | \boldsymbol{\Lambda}_k, \boldsymbol{\Psi}_k, \boldsymbol{\Omega}_k, \boldsymbol{Y}_k^{obs}] \\ & \triangleq N[(\boldsymbol{\Sigma}_0^{-1} + n_k \boldsymbol{\Psi}_k^{-1})^{-1} (n_k \boldsymbol{\Psi}_k^{-1} \bar{\boldsymbol{Y}}_k + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{u}_0), (\boldsymbol{\Sigma}_0^{-1} + n_k \boldsymbol{\Psi}_k^{-1})^{-1}], \end{aligned}$$

and for $j = 1, \cdots, p$,

$$\begin{aligned} & [\boldsymbol{\Lambda}_{kj} | u_{kj}, \psi_{kj}, \boldsymbol{\Omega}_k, \boldsymbol{Y}_k^{obs}] \triangleq N[\boldsymbol{\Lambda}_{ykj}, \psi_{kj} \boldsymbol{\Sigma}_{ykj}], \\ & [\psi_{kj}^{-1} | u_{kj}, \boldsymbol{\Omega}_k, \boldsymbol{Y}_k^{obs}] \triangleq Gamma[n/2 + \alpha_{0\epsilon k}, \beta_{\epsilon kj}], \end{aligned}$$

where $\bar{\boldsymbol{Y}}_k = \sum_{i:\mathrm{W}_i = k} \frac{\boldsymbol{y}_i^{obs} - \boldsymbol{\Lambda}_k \boldsymbol{\omega}_i}{n_k}$, with $\sum_{i:\mathrm{W}_i = k}$ be the summation with respect to those i such that $\mathrm{w}_i = k$, and $n_k = \sum_{i=1}^{n} I(\mathrm{w}_i = k)$, $\boldsymbol{\Lambda}_{ykj} = \boldsymbol{\Sigma}_{ykj} (\boldsymbol{H}_{0ykj}^{-1} \boldsymbol{\Lambda}_{0kj} +$

$\boldsymbol{\Omega}_k\tilde{\boldsymbol{Y}}_{kj}^T$), $\boldsymbol{\Sigma}_{ykj} = (\boldsymbol{H}_{0ykj}^{-1} + \boldsymbol{\Omega}_k\boldsymbol{\Omega}_k^T)^{-1}$, $\tilde{\boldsymbol{Y}}_{kj} = \sum_{i:\mathrm{W}_i=k}(y_{ij} - u_{kj})\boldsymbol{\omega}_i$ and $\beta_{\epsilon kj} = \beta_{0\epsilon k} + [\tilde{\boldsymbol{Y}}_{kj}\tilde{\boldsymbol{Y}}_{kj}^T - \boldsymbol{\Lambda}_{ykj}^T\boldsymbol{\Sigma}_{ykj}^{-1}\boldsymbol{\Lambda}_{ykj} + \boldsymbol{\Lambda}_{0kj}^T\boldsymbol{H}_{0ykj}^{-1}\boldsymbol{\Lambda}_{0kj}]$.

For the parameters involved in $\boldsymbol{\theta}_{\omega k}$, we consider the following prior distributions:

$$p(\boldsymbol{\Phi}_k) \triangleq IW_{q_2}(\boldsymbol{R}_0, \rho_0),$$

and for $j = 1, \cdots, q_1$,

$$p(\boldsymbol{\Lambda}_{\omega kj}|\psi_{\delta kj}) \triangleq N[\boldsymbol{\Lambda}_{0\omega kj}, \psi_{\delta kj}\boldsymbol{H}_{0\omega kj}],\ p(\psi_{\delta kj}^{-1}) \triangleq Gamma(\alpha_{0\delta k}, \beta_{0\delta k}),$$

where $\boldsymbol{\Lambda}_{\omega k} = (\boldsymbol{\Pi}_k, \boldsymbol{\Gamma}_k)$, $\boldsymbol{\Lambda}_{\omega kj}$ is the jth row of $\boldsymbol{\Lambda}_{\omega k}$, and $\psi_{\delta kj}$ is the jth element of $\boldsymbol{\Psi}_{\delta k}$; $\boldsymbol{\Lambda}_{0\omega kj}, \alpha_{0\delta k}, \beta_{0\delta k}$, ρ_0, and the positive definite matrix $\boldsymbol{R}_0$ and $\boldsymbol{H}_{0\omega kj}$ are all given hyperparameters. With these prior distributions, we can get the posterior distributions of them as below by some derivations.

$$[\boldsymbol{\Phi}_k|\boldsymbol{\Omega}_{2k}] \triangleq IW_{q_2}[\boldsymbol{\Omega}_{2k}\boldsymbol{\Omega}_{2k}^T + \boldsymbol{R}_0^{-1}, n_k + \rho_0],$$

and for $j = 1, \cdots, q_1$,

$$[\boldsymbol{\Lambda}_{\omega kj}|\psi_{\delta kj}^{-1}, \boldsymbol{\Omega}_k, \boldsymbol{Y}_k^{obs}] \triangleq N[\boldsymbol{\Lambda}_{\delta kj}, \psi_{\delta kj}\boldsymbol{\Sigma}_{\omega kj}],$$
$$[\psi_{\delta kj}^{-1}|\boldsymbol{\Omega}_k, \boldsymbol{Y}_k^{obs}] \triangleq Gamma[n/2 + \alpha_{0\delta k}, \beta_{\delta kj}],$$

where $\boldsymbol{\Lambda}_{\delta kj} = \boldsymbol{\Sigma}_{\omega kj}(\boldsymbol{H}_{0\omega kj}^{-1}\boldsymbol{\Lambda}_{0\omega kj} + \boldsymbol{\Omega}_k\boldsymbol{\Omega}_{1kj}^T)$, $\boldsymbol{\Sigma}_{\omega kj} = (\boldsymbol{H}_{0\omega kj}^{-1} + \boldsymbol{\Omega}_k\boldsymbol{\Omega}_k^T)^{-1}$, and $\beta_{\delta kj} = \beta_{0\delta k} + [\boldsymbol{\Omega}_{1kj}\boldsymbol{\Omega}_{1kj}^T - \boldsymbol{\Lambda}_{\delta kj}^T\boldsymbol{\Sigma}_{\omega kj}^{-1}\boldsymbol{\Lambda}_{\delta kj} + \boldsymbol{\Lambda}_{0\omega kj}^T\boldsymbol{H}_{0\omega kj}^{-1}\boldsymbol{\Lambda}_{0\omega kj}]$. The above posterior distributions are the familiar normal, gamma, and wishart distributions. Generating observations from them are simple.

5.3.4 Calibration Distribution

To define the calibration distribution, let M_c denote the candidate model under consideration, and M_t denote the true model. Furthermore, let $L_v(\boldsymbol{Y}^{obs}, M_c)$ denote the L_v measure of the candidate model M_c, and $L_v(\boldsymbol{Y}^{obs}, M_t)$ denote the L_v measure of the true model M_t. Then given v, the difference of L_v measures between the models M_c and M_t is defined as

$$D_v(\boldsymbol{Y}^{obs}, M_c) \equiv L_v(\boldsymbol{Y}^{obs}, M_c) - L_v(\boldsymbol{Y}^{obs}, M_t). \tag{5.23}$$

Similar as discussed in previous chapters, to calibrate the L_v measure, we need to construct the marginal distribution of $D_v(\boldsymbol{Y}^{obs}, M_c)$, computed with respect to the prior predictive distribution of $\boldsymbol{Y}^{obs}$ under the true model M_t:

$$p_t(\boldsymbol{Y}^{obs}) = \int p(\boldsymbol{Y}^{obs}|\boldsymbol{\theta}, \boldsymbol{\Omega}, M_t)p(\boldsymbol{\theta}|M_t)d\boldsymbol{\theta}, \tag{5.24}$$

where $p(\boldsymbol{\theta}|M_t)$ is proper prior distribution for $\boldsymbol{\theta}$ under the true model M_t. We denote the calibration distribution of the candidate model M_c by

$$PL_c \equiv p(D_v(\boldsymbol{Y}^{obs}, M_c)). \tag{5.25}$$

After obtaining the calibration distribution PL_c, the mean $\mu_v(M_c)$, the standard deviation $SD_v(M_c)$, and the highest probability density (HPD) interval of PL_c can be obtained. As we can get a closed form of neither the L_v measure, nor the calibration distribution, MCMC methods are used to estimated them. Details can be found in Chapter 2.

5.4 A Simulation Study

Suppose the observations $\boldsymbol{Y}^{obs}$ are generated from a mixture SEM with two components defined by equations (5.1), (5.2), and (5.3). The loading matrices $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are given by

$$\boldsymbol{\Lambda}_1^T = \boldsymbol{\Lambda}_2^T = \begin{pmatrix} 1.0^* & \lambda_{21} & \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{52} & \lambda_{62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{83} & \lambda_{93} \end{pmatrix},$$

where the elements with asterisks are fixed at preassigned values. True values of the parameters in this model are given by: $\pi_1 = 0.5, \pi_2 = 0.5, \boldsymbol{u}_1 = 0 \times \mathbf{1}_9, \boldsymbol{u}_2 = 2 \times \mathbf{1}_9$, where $\mathbf{1}_9$ is a 9-dimensional vector with all elements being 1; $\lambda_{21} = \lambda_{31} = \lambda_{52} = \lambda_{62} = \lambda_{83} = \lambda_{93} = 0.8$, $\boldsymbol{\Gamma}_1 = (0.5, 0.5), \boldsymbol{\Gamma}_2 = (0.5, -0.5)$, $\boldsymbol{\Psi}_1 = \text{diag}(0.64), \boldsymbol{\Psi}_2 = \text{diag}(0.36), \psi_{\delta 1} = 0.5, \psi_{\delta 2} = 1.0$, and $\phi_{k,11} = \phi_{k,22} = 1.0, \phi_{k,12} = 0.3, \phi_{k,12} = -0.3$ for $k = 1, 2$. For convenience, we denote this two-component model by M_2. To evaluate the performance

of the L_v measure, three SEMs with different components will be considered as competing models. Let M_1, M_3, and M_4 be mixture SEMs with one, three, and four components, respectively. In this simulation study, two samples with $n = 300$ and 600 are considered, and 100 replications are conducted to estimate the calibration distribution for each sample.

To give a rough analysis of the impact of prior inputs on model selection, two types of prior inputs, Prior I and Prior II, are considered.

Prior I: The means of the normal distributions are taken as the true values of the corresponding parameters, and the covariance matrices are taken as the identity matrices with corresponding dimensions; $\boldsymbol{R}_0$ equals 4 times of the identity matrix, and $\rho_0 = 7$ in the Wishart distribution; $\alpha_{0\epsilon k} = \alpha_{0\delta k} = 5$ and $\beta_{0\epsilon k} = \beta_{0\delta k} = 8$.

Prior II: The means of the normal distributions are taken as zero, the covariance matrices are equal to four times of the identity matrices with appropriate dimensions; $\boldsymbol{R}_0$ equals 8 times of the identity matrix, and $\rho_0 = 8$ in the Wishart distribution; $\alpha_{0\epsilon k} = \alpha_{0\delta k} = 10$ and $\beta_{0\epsilon k} = \beta_{0\delta k} = 10$.

L_v measure for model selection

The mean values of the L_v measures and the summaries of the calibration distributions are given in Table 5.1, where mean($L_{0.5}$) denotes the mean value of the L_v measure with $v = 0.5$, $\mu_{0.5}$, $SD_{0.5}$, and 95% HPD denote the mean, the standard deviation, and the 95% HPD interval of the calibration distribution of the corresponding model with $v = 0.5$, respectively. From this table, we see that $\mu_{0.5}$ and the 95% HPD interval corresponding to M_1 are far apart from zero, so M_2 performs much better than M_1 under each case. However, for M_3 and M_4, $\mu_{0.5}$ values are relatively close to zero, and zero is included in the 95% HPD intervals. Therefore, the performances of M_2, M_3, and M_4 are similar. Based on the parsimonious principle, the simpler model M_2 is selected. The calibration distributions under each given type of prior inputs and sample size are presented in Figure 5.1(a)

to Figure 5.2(b). From these figures, the calibration distributions of M_3 and M_4 are centered at zero, while the calibration distribution of M_1 is far apart from zero under each case. This gives the same conclusion as above.

Bayes factor for model selection

In computing Bayes factor through path sampling, finding a good linking model to link two competing models is an important step. An illustrate example is given as follows. Consider two models M_2 and M_3 in this simulation study. These two models can be written as:

$$M_2 : f(\boldsymbol{y}_i^{obs}|\boldsymbol{\theta}, \boldsymbol{\pi}^*) = \sum_{k=1}^{2} \pi_k^* f_k(\boldsymbol{y}_i^{obs}|\boldsymbol{u}_k, \boldsymbol{\Sigma}_k),$$

$$M_3 : (\boldsymbol{y}_i^{obs}|\boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{k=1}^{3} \pi_k f_k(\boldsymbol{y}_i^{obs}|\boldsymbol{u}_k, \boldsymbol{\Sigma}_k).$$

To apply path sampling in computing $\log B_{32}$, they are linked up by a path $t \in [0, 1]$ as follows:

$$\begin{aligned} M_t : f(\boldsymbol{y}_i^{obs}|\boldsymbol{\theta}, \boldsymbol{\pi}, t) &= [\pi_1 + (1-t)a_1\pi_3] f_1(\boldsymbol{y}_i^{obs}|\boldsymbol{u}_k, \boldsymbol{\Sigma}_k) \\ &\quad + [\pi_2 + (1-t)a_2\pi_3] f_2(\boldsymbol{y}_i^{obs}|\boldsymbol{u}_k, \boldsymbol{\Sigma}_k) + t\pi_3 f_3(\boldsymbol{y}_i^{obs}|\boldsymbol{u}_k, \boldsymbol{\Sigma}_k), \end{aligned}$$

where $a_1 + a_2 = 1$. Clearly, when $t = 1$, M_t reduces to M_3; when $t = 0$, M_t reduces to M_2 with $\pi_k^* = \pi_k + a_k\pi_3, k = 1, 2$. A natural choice is $a_k = 1/2, k = 1, 2$. With this linking model, logarithm Bayes factor can be estimated (see Chapter 2).

The logarithm Bayes factors in this simulation study are given in Table 5.2. From this table, the same model M_2 is selected under each given prior inputs.

5.5 A Real Example

In this section, a small portion of the ICPSR data set collected in the project WORLD VALUES SURVEY 1981-1984 AND 1990-1993 (World

Value Study Group, ICPSR Version) is analyzed. As an illustration of the proposed method, only the data obtained from United Kingdom with a sample size 1484 are used. Eight variables (V116, V117, V180, V132, V96, V255,V254, and V252) in the original data set that related to respondents 'job' and 'homelife' are taken as manifest variables in $\boldsymbol{y}$. A description of these variables in the questionnaire is given in Appendix. After deleting the cases with missing entries, the sample size is 819. In fact, this data set has been analyzed by Lee (2007), they proposed mixture SEMs for this data set and applied Bayes factor to model selection. In this chapter, five mixture SEMs as discussed in Lee (2007) are considered, and L_v measure is used for model selection. Specifically, these models are denoted by M_k for $k = 1, \cdots, 5$, where k denotes the number of components in M_k. For each component in these models, there are three latent variables which can be roughly interpreted as 'job satisfaction η', 'homelife, ξ_1' and 'job attitude, ξ_2'. Then these models can be defined by equations (5.1), (5.2), and (5.3), and the specification of the parameter matrices in the model formulation are given by $\boldsymbol{\Pi}_k = 0, \boldsymbol{\Psi}_{\delta k} = \psi_{\delta k}, \boldsymbol{\Gamma}_k = (\gamma_{k,1}, \gamma_{k,2})$, $\boldsymbol{\Psi}_k = \text{diag}(\psi_{k1}, \cdots, \psi_{k8})$ and

$$\boldsymbol{\Lambda}_k^T = \begin{bmatrix} 1.0^* & \lambda_{21} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 1.0^* & \lambda_{42} & \lambda_{52} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{73} & \lambda_{83} \end{bmatrix}, \boldsymbol{\Phi}_k = \begin{bmatrix} \phi_{k,11} & \phi_{k,12} \\ \phi_{k,21} & \phi_{k,22} \end{bmatrix}.$$

The elements in $\boldsymbol{\Lambda}_k$ with an asterisk are fixed for identification. To calculate L_v measure, MCMC methods are used. In this example, the inputs for the prior distribution of the unknown parameters are taken such that the means of the normal distributions are taken as the estimates of the corresponding parameters obtained by using noninformative prior inputs, the covariance matrices of the normal distributions are taken as the identity matrices with appropriate dimension; $\boldsymbol{R}_0$ equals 5 times of the identity matrix, and $\rho = 5$ in the Wishart distribution; $\alpha_{0\epsilon k} = \alpha_{0\delta k} = 6$ and $\beta_{0\epsilon k} = \beta_{0\delta k} = 6$ in the

gamma distributions. The L_v measure for these five models are:

$$L_{0.5}(\boldsymbol{Y}^{obs}, M_1) = 34547.15,\ L_{0.5}(\boldsymbol{Y}^{obs}, M_2) = 32368.39,$$
$$L_{0.5}(\boldsymbol{Y}^{obs}, M_3) = 30323.69,\ L_{0.5}(\boldsymbol{Y}^{obs}, M_4) = 30680.77,$$
$$L_{0.5}(\boldsymbol{Y}^{obs}, M_5) = 31061.87.$$

From the results, the mixture model with three components has the smallest value of L_v measure. Therefore, M_3 is selected based on L_v measure. However, as discussed before, only considering the model with the smallest criterion value may be misleading, thus calibration distribution is considered. To estimate the calibration distribution, 100 pseudo data sets are generated from $p(\boldsymbol{Y}|\boldsymbol{\theta}, M_3)p(\boldsymbol{\theta})$, where $p(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$ with proper prior inputs. Through MCMC method, a sample of $D_{0.5}(M_k)(k = 1, 2, 4, 5)$ can be obtained, and the calibration distributions of the four models are estimated with kernel density estimation methods. The calibration summaries are given in Table 5.3. From this table, $\mu_{0.5}(M_1)$ and $\mu_{0.5}(M_2)$ are larger than zero, and the 95% HPD intervals of these two models are far apart from zero. Therefore, M_3 performs much better than M_1 and M_2. However, $\mu_{0.5}(M_4)$ and $\mu_{0.5}(M_5)$ are less than zero but relatively close to zero, and zero is included in the 95% HPD intervals of these two models. Hence, the performances of M_3, M_4 and M_5 are similar. According to the parsimonious principle, the simpler model, M_3, is selected. The calibration distributions for M_1, M_2, M_4, and M_5 are presented in Figure 5.3. From this figure, M_3 performs significantly better than M_1 and M_2, while the performances of M_3, M_4, and M_5 are similar. Thus the same conclusion as above can be obtained. This data set was analyzed by Lee (2007, chp. 11). They conducted model selection with Bayes factor. According to their result, M_3 is also selected as the best model among the five competing models.

5.6 Discussion

In this chapter, L_v measure and the calibration distributions are applied to model selection of finite mixture SEMs. As for a mixture model, no information about which component the individual belongs to, we consider not only the goodness-of-fit of the model, but also the classification accuracy in the L_v measure. Based on the numerical studies, our proposed L_v measure performs satisfactorily in model selection of mixture SEMs. The involved computation is easy and fast.

Table 5.1: Mean values of the L_v measure and calibration and calibration summaries for simulation study

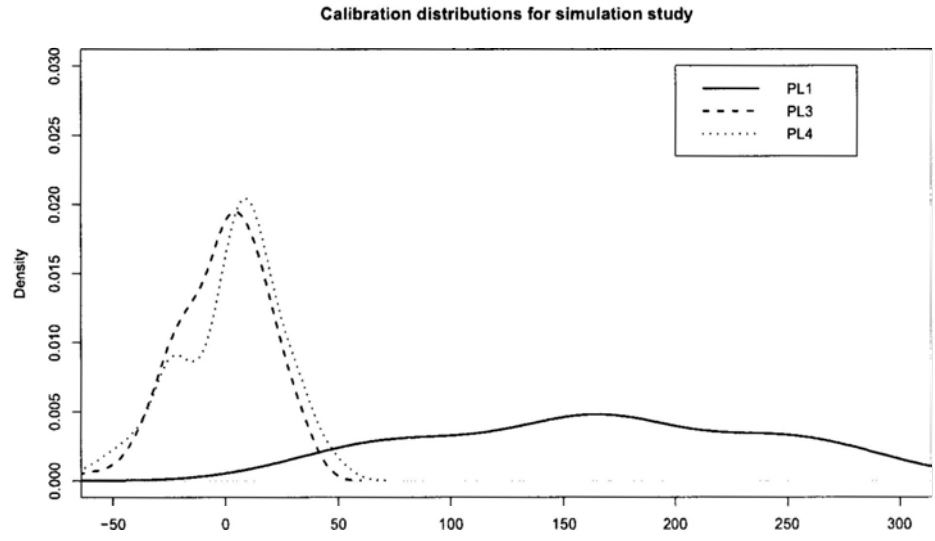
Prior	Sample size	Model	mean($L_{0.5}$)	$\mu_{0.5}$	$SD_{0.5}$	95% HPD
Type I	n=300	M_2	3157.436	-	-	-
		M_1	3325.981	168.545	78.274	(39.502, 307.358)
		M_3	3155.765	-1.671	21.408	(-44.355, 33.198)
		M_4	3159.633	2.198	22.583	(-45.667, 42.495)
	n=600	M_2	6246.697	-	-	-
		M_1	6746.575	499.877	112.245	(264.720, 692.098)
		M_3	6240.849	-5.848	9.705	(-34.112, 9.453)
		M_4	6238.067	-8.630	13.690	(-41.791, 12.232)
Type II	n=300	M_2	3735.657	-	-	-
		M_1	3822.117	86.460	55.856	(1.987, 197.129)
		M_3	3742.335	6.678	5.487	(-4.698, 17.288)
		M_4	3748.650	12.993	7.014	(1.658, 25.841)
	n=600	M_2	6898.080	-	-	-
		M_1	7294.296	396.217	109.127	(150.538, 574.789)
		M_3	6903.009	4.930	7.404	(-9.959, 19.057)
		M_4	6908.377	10.297	7.366	(-4.014, 24.985)

Table 5.2: Logarithm Bayes factor for simulation study

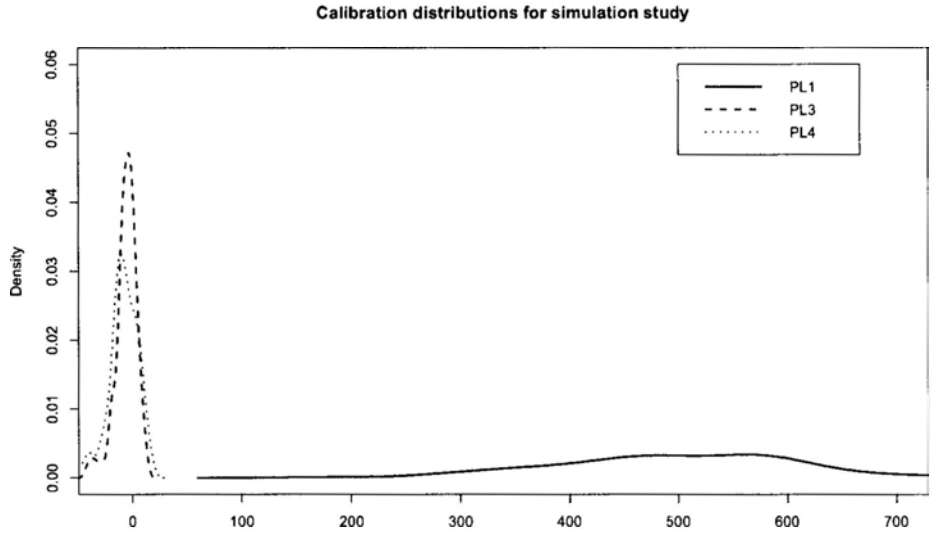
Prior	Logarithm Bayes factor	n=300	n=600
Type I	$\log B_{21}$	116.1152	272.7424
	$\log B_{32}$	-0.0590	-1.1253
	$\log B_{43}$	-1.1631	-1.2890
Type II	$\log B_{21}$	76.8860	238.6673
	$\log B_{32}$	0.4207	-1.1253
	$\log B_{43}$	-0.4012	-0.2890

Table 5.3: Calibration summaries for real example

Model	$\mu_{0.5}$	$SD_{0.5}$	95% HPD
M_1	3861.841	462.727	(2924.941, 4575.270)
M_2	1190.338	283.321	(643.688, 1708.400)
M_4	-4.170	157.924	(-384.711, 236.816)
M_5	-18.342	166.512	(-381.604, 252.348)

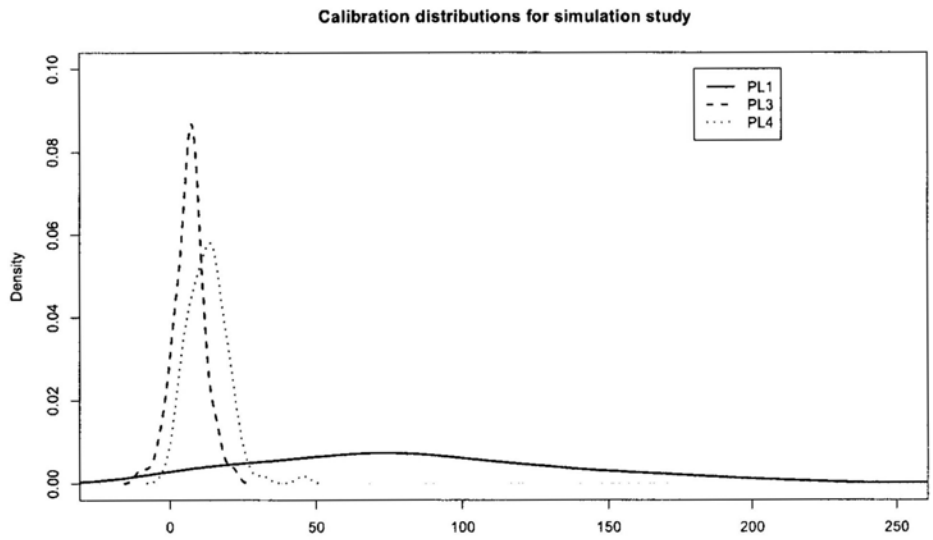


(a) n=300

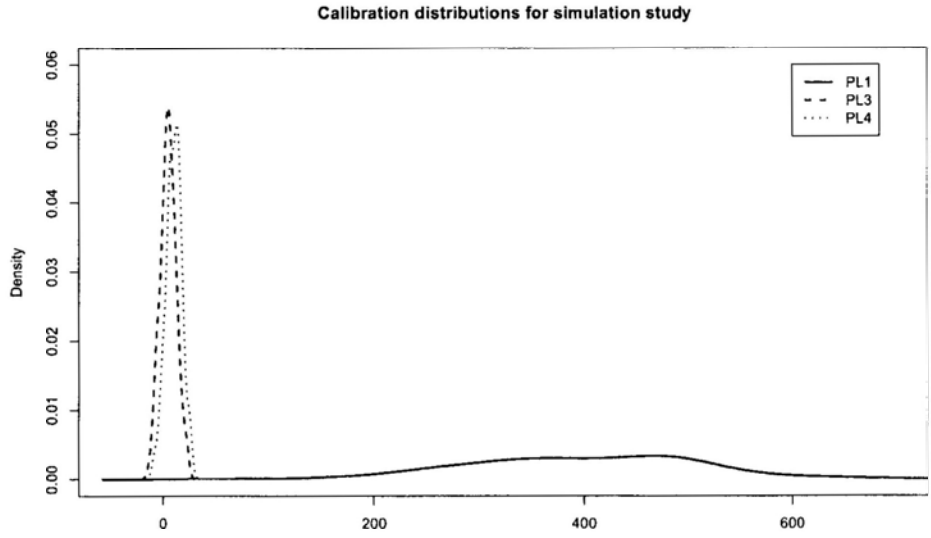


(b) n=600

Figure 5.1: Calibration distributions for simulation study under Prior I



(a) n=300



(b) n=600

Figure 5.2: Calibration distributions for simulation study under Prior II

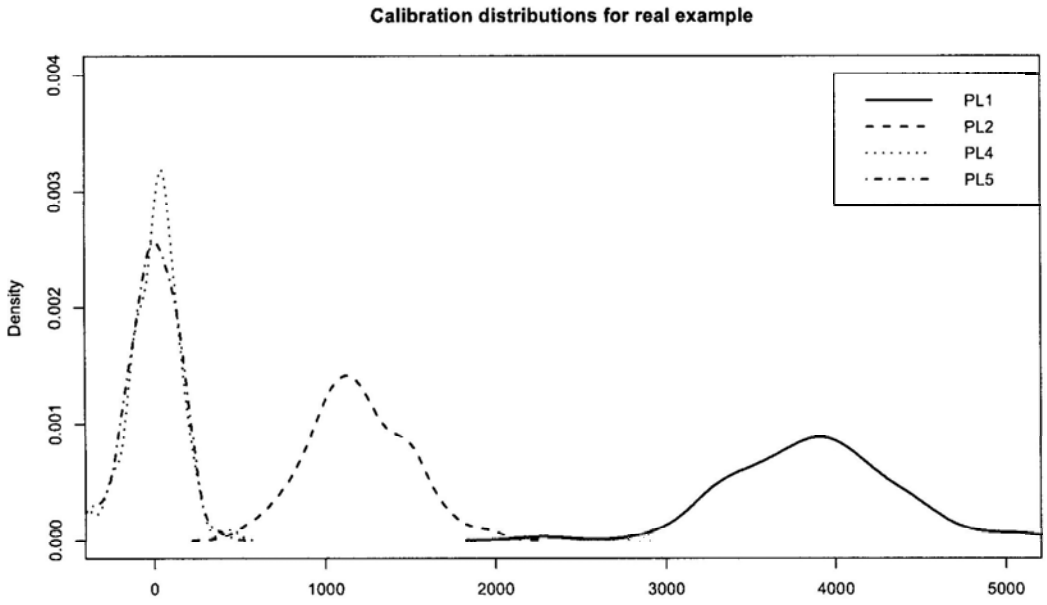


Figure 5.3: Real Example

Chapter 6

Conclusions and Further Developments

6.1 Conclusions

Structural equation models (SEMs) are widely used in behavioral, educational, medical and social sciences. In this thesis, we focus on model selection for four different kinds of SEMs. In Chapter 2, the L_v measure is used for model selection of nonlinear SEMs with continuous data, and from the illustrative examples, the performance of the L_v measure for nonlinear SEMs is pretty good. Discrete responses are often encountered in structural equation modeling due to the questionnaire design. Therefore, there is a need to propose SEMs with categorical data. Motivated by this problem, we further apply the L_v measure to nonlinear SEMs with mixed continuous and ordinal categorical data in Chapter 3. Moreover, in the analysis of data from behavioral, educational, medical and social science, heterogeneity is inevitable. To deal with heterogeneous data, multi-level SEMs and mixture SEMs are developed. Searching for a good model selection method for these two kinds of models becomes an important issue. Thus, in Chapter 4, we propose the L_v measure for model selection of multi-level SEMs. Specifically, a two-level SEMs is presented as an example. In Chapter 5, the L_v measure is proposed for model selection of finite mixture SEMs.

Through all these chapters, MCMC methods are used to calculate the L_v measure, and kernel density estimation method is applied to estimate the calibration distributions. According to our analysis, the L_v measure can be easily calculated after obtaining the Beyesian estimation of the unknown parameters and latent variables. We conduct simulation studies to demonstrate the proposed methods in each of these chapters. For illustration, the developed methodologies are also applied to analyze some real data sets. Moreover, results of model selection based on the L_v measure are compared with those based on Bayes factor (Kass and Raftery, 1995) and deviance information criterion (DIC) (Spiegelhalter et al., 2002).

6.2 Discussion and Further Developments

From the simulation studies, only considering the smallest value of the L_v measure may lead to misleading results in some circumstances. Especially when comparing two nested models, the full model may have a slightly smaller value than the alternative model. Therefore, calibration distribution and a parsimonious method are necessary. In this thesis, the mean, the standard deviance, and the HPD interval of the calibration distributions are considered. However, there are some other summaries that may be helpful for making decision about model selection. For example, the p-value of the calibration distribution.

In this thesis, we take the value of v in L_v measure as 0.5, which gives equal weights to the departure of the 'guess' to the future values and to the observed values. However, as discussed in Ibrahim et al. (2001), changing the value of v gives more flexibility to the L_v measure in model selection. Therefore, selecting the value of v in model selection with the L_v measure should be an important issue. We can try to find the optimal v in comparing different kinds of models.

Moreover, according to the original definition of the L_v measure given

by Gelfand and Ghosh (1998), different criterion can be obtained by using different loss function. In all the chapters, we applied the expected squared Euclidean distance (Ibrahim and Laud, 1994) as the loss function. That is because the computation of the L_v measure using this loss function is quite simple. However, as we discussed before, calibration distributions are important in using such a measure. As computing the calibration distributions is time consuming, and it requires proper prior distributions of the unknown parameters, there is a need to find a more efficient criterion which can produce better result. For example, Gelfand and Ghosh (1998) proposed a deviance version of the criterion by using log scoring loss notion. We can try to use this criterion as a statistic for model selection of SEMs, or we can find other criterions by using different loss functions.

Based on our study, the L_v measure can be used to the problem of model selection for the hierarchical models with latent variables. In this thesis, we mainly consider the SEMs in which the response variables are normally distributed. However, in the application of SEMs, the variables may come from an exponential family (Song and Lee, 2007; Lee, 2007). Therefore, there is a need to propose the L_v measure for SEMs with data from exponential family. Furthermore, incomplete or missing data are commonly encountered in practical applications. Basically, there are two kinds of missing data, ignorable missing and nonignorable missing. Rubin (1976) provided a classification system for missing data mechanisms and argued that missing data can be ignored under two conditions: missing completely at random (MCAR) and missing at random (MAR). Missing data are nonignorable in the sense that the missing values depend on themselves (Little and Rubin, 1987). Thus, finding an appropriate missing mechanism based on the L_v measure is also an important issue in structural equation modeling.

References

Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle.* in Second International Symposium on Information Theory.

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, 16:3–14.

Ansari, A., Jedidi, K., and Dube, L. (2002). Information theory and an extension of the maximum likelihood principle. *Psychometrika*, 67:49–78.

Bayarri, M. J. and Berger, J. O. (1999). Quantifying surprise in the data and model verification. *In Bayesian Statistics 6*, pages 53–82.

Box, G. E. P. (1980). Sampling and bayes inference in scientific modelling and robustness (with discussion). *Journal of the Royal Statistical Society: Series A*, 143:383–430.

Cai, J. H., Song, X. Y., and Hser, Y. I. (2010). A Bayesian analysis of mixture structural equation models with nonignorable missing responses and covariates. *Statistics in Medicine, in press.*

Celeux, G., Forbes, F., Robert, C., and Titterington, M. (2006). Deviance information criteria for missing data. *Bayesian Analysis*, 1(4):651–674.

Chen, M. H., Dey, D. K., and Ibrahim, J. G. (2004). Bayesian criterion based model assessment for categorical data. *Biometrika*, 91:45–63.

Chen, M.-H. and Shao, Q.-M. (1999). Monte carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8:69–92.

DiCiccio, T. J., Kass, R. E., Raftery, A., and Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, 92:903–915.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B*, 56:363–375.

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B*, 62:355–366.

Frühwirth-Schnatter, S. (2001). Markov chain monte carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96:194–208.

Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85:1–11.

Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13:163–185.

Gelman, A., Meng, X. L., and Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistic Sinica*, 6:733–807.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Interlligence*, 6:721–741.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their application. *Biometrika*, 57:97–109.

Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, 13:795–800.

Ibrahim, J. G., Chen, M., and Sinha, D. (2001). Criterion-based methods for Bayesian model assessment. *Statistica Sinica*, 11:419–443.

Ibrahim, J. G. and Laud, P. W. (1994). A predictive approach to the analysis of designed experiments. *Journal of the American Statistical Association*, 89:309–319.

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

Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International: Hove and London.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90:773–795.

Lee, S. Y. (2007). *Structural Equation Modelling: A Bayesian Approach*. Wiley.

Lee, S. Y. and Shi, J. Q. (2001). Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics*, 57:787–794.

Lee, S. Y. and Song, X. Y. (2001). Hypothesis testing and model comparison in two-level structural equation models. *Multivariate Behavioral Research*, 66:639–655.

Lee, S. Y. and Song, X. Y. (2002). Bayesian selection on the number of factors in a factor analysis model. *Behaviormetrika*, 29:23–39.

Lee, S. Y. and Song, X. Y. (2003a). Bayesian model selection for mixtures of structural equation models with an unknown number of components. *British Journal of Mathematical and Statistical Psychology*, 56:145–165.

Lee, S. Y. and Song, X. Y. (2003b). Model comparison of nonlinear structural equation models with fixed covariates. *Psychometrika*, 68:27–47.

Lee, S. Y. and Song, X. Y. (2004). Bayesian model comparison of nonlinear latent variable models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, 57:131–150.

Lee, S. Y. and Song, X. Y. (2005). Maximum likelihood analysis of a two-level nonlinear structural equation model with fixed covariates. *Journal of Educational and Behavioral Statistics*, 30:1–26.

Lee, S. Y., Song, X. Y., and Cai, J. H. (2010). A Bayesian approach for nonlinear structural equation models with dichotomous variables using logit and probit links. *Structural Equation Modeling - A Multidisciplinary Journal*.

Lee, S. Y. and Zhu, H. T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous dta. *British Journal of Mathematical and Statistical Psychology*, 53:209–232.

Lindsay, B. G. and Basak, P. (1993). Multivariate normal mixtures: A fast consistent method of moments. *Journal of the American Statistical Association*, 88:468–476.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

McDonald, R. P. and Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *The British Journal of Mathematical and Statistical Psychology*, 42:215–232.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 21:1087–1091.

Power, M., Bullingen, M., and Hazper, A. (1999). The world health organization whoqol-100: test of the universality of quality of life in 15 different cultural groups worldwide. *Health Psychology*, 18:495–182.

Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the em algorithm. *SIAM Review*, 26:195–239.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society, Series B*, 59:731–792.

Robert, C. (1996). Mixtures of distributions: Inference and estimation. *In Practical Markov Chain Monte Carlo*, (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter), London: Chapman and Hall:163–188.

Robert, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7:110–120.

Roeder, K. and Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, 92:894–902.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63:581–592.

Schines, R., Hoijtink, H., and Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64:37–52.

Schwarz, G. (1978). Estimating the dimension of a model. *Annal of Statistics*, 6:461–464.

Scott, D. W. (1992). *Multivariate Density Estimation. Theory, Practice and Visualization.* New York: Wiley.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B*, 53:683–690.

Silverman, B. W. (1986). *Density Estimation for statistics and data analysis.* London: Chapman and Hall.

Song, X. Y. and Lee, S. Y. (2004). Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, 57:29–52.

Song, X. Y. and Lee, S. Y. (2007). Bayesian analysis of latent variable models with non-ignorable missing outcomes from exponential family. *Statistics in Medicine*, 26:681–693.

Song, X. Y. and Lee, S. Y. (2008). A Bayesian approach for analyzing hierarchical data with missing outcomes through structural equation models. *Structural Equation Modeling - A Multidisciplinary Journal*, 15:272–300.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). *WinBUGS user manual, version 1.4.* MRC Biostatistics Unit, Cambridge, UK.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, 64:583–639.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions.* John Wiley, New York.

Zhang, W. and Lee, S. Y. (2001). Asymptotic theory of two-level structural equation models with constraints. *Statistica Sinica*, 11:135–145.

Zhu, H. T. and Lee, S. Y. (2001). A Bayesian analysis of finite mixtures in the lisrel model. *Psychometrika*, 66:133–152.

Appendix

V62 (*Religious beliefs*): 1, 2, 3, 4, 5.

V96 (*All things considered, how satisfied are you with your life as a whole these days?*): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (from "Dissatisfied" to "Satisfied").

V116 (*Overall, how satisfied or dissatisfied are you with your job?*): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (from "Dissatisfied" to "Satisfied").

V117 (*How free are you to make decisions in your job?*): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (from "Not at all" to "A great deal").

V132 (*How satisfied are you with the financial situation of your household?*): 1, 2, 3 ,4 ,5, 6 ,7 ,8 ,9 ,10 (from "Dissatisfied" to "Satisfied").

V179 (*How often do you pray to God outside of religious services? Would you say ...*): 1 (Often), 2 (Sometimes), 3 (Hardly ever), 4 (Only in times of crisis), 5 (Never).

V180 (*Overall, how satisfied or dissatisfied are you with your home life?*): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (from "Dissatisfied" to "Satisfied").

V252 (*Individual should take more respectively for providing for themselves*)1, 2, 3, 4, 5 ,6 ,7 ,8 ,9 ,10 (*The state should take more responsibility to ensure that everyone is provided for*).

V254 (*Competition is good. It stimulates people to work hard and develop new ideas*)1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (*It Competition is harmful. brings out the worst in peroples*).

V255 (*In the long run, hard work usually brings a better life*) 1, 2, 3, 4, 5 ,6 ,7, 8, 9, 10 (*Hard work doesn't generally brings success–it's more a matter of luck and connections*).