A Bayesian Criterion-based Statistic for Model Selection of Structural Equation Models with Ordered Categorical data

Yunxian Li and Aijing Yang

***Abstract*—In this paper, a Bayesian criterion-based method called the Lv measure, as well as its calibration distribution, is introduced and applied to model selection of structural equation models with ordered categorical data. A simulation study is presented to illustrate the satisfactory performance of the Lv measure in model selection. A quality of life data is analyzed for illustration.**

***Index Terms*—ordered categorical data; Lv measure; Model selection; MCMC algorithm.**

I. Introduction

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 [1] for example. An important issue in the application of SEMs with ordered categorical data is to optimize the given model. In this paper, we treat this problem as model selection, and apply a Bayesian criterion-based method, which is called the Lv measure [2], to model selection. The Lv measure involves two components, the first one is related with the reliability, and the other one measures the discrepancy between the predictions and the observations. Thus, the model with the smallest Lv measure is considered as the optimal model. We also considered the calibration distribution of the Lv measure, which will allow us to compare two competing models formally. The Bayesian approach, together with MCMC algorithms, is used to estimate the latent variables and unknown parameters, and to compute the Lv measure for SEMs with ordered categorical data.

The remainder of this paper is divided into four sections. In Section II, model selection of nonlinear SEMs with ordered categorical data will be discussed. In Section III, a simulation study is presented to demonstrate the performance of the Lv measure. In Section IV, a real example is analyzed. A discussion is given in Section V.

Manuscript received April 1, 2011; revised May 25, 2011.

Yunxian Li is with the School of Finance, Yunnan University of Finance and Economics, Kunming, Yunnan, China. (phone: 86-15288336639;fax: 86-871-5113384; email: cloudlicu@gmail.com).

Aijun Yang is with the School of Finance, Nanjing Audit University, Nanjing, Jiangsu, China. (e-mail: ajyang81@hotmail.com).

II. Model Selection of Structural Equation Models with Ordered Categorical Data

A. Nonlinear structural equation model with ordered categorical data

Let $\boldsymbol{y}_i$ ($i=1,\cdots,n$) be a $p\times 1$ random vector of observed variables, and $\boldsymbol{Y}=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)$. The measurement and structural equations of the structural equation model is defined by

$$\begin{aligned} M:\ \boldsymbol{y}_i &= \boldsymbol{u}+\boldsymbol{\Lambda}\boldsymbol{\omega}_i+\boldsymbol{\varepsilon}_i, \\ \boldsymbol{\eta}_i &= \boldsymbol{\Pi}\boldsymbol{\eta}_i+\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i)+\boldsymbol{\delta}_i, \end{aligned} \tag{1}$$

where $\boldsymbol{u}$ is a $p\times 1$ mean vector; $\boldsymbol{\omega}_i$ is a $q\times 1$ vector of latent variables; $\boldsymbol{\varepsilon}_i$ is a $p\times 1$ random vector of error terms, and is independent of $\boldsymbol{\omega}_i$; $\boldsymbol{\omega}_i=(\boldsymbol{\eta}_i^T,\boldsymbol{\xi}_i^T)^T$, in which $\boldsymbol{\eta}_i(q_1\times 1)$ and $\boldsymbol{\xi}_i(q_2\times 1)$ are vectors of endogenous and exogenous latent variables, 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 vector-valued function with differentiable functions $f_1(\cdot),\cdots,f_r(\cdot)$, 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 $i=1,\cdots,n$,

$$\boldsymbol{\varepsilon}_i\sim N[\boldsymbol{0},\boldsymbol{\Psi}_\varepsilon],\ \boldsymbol{\delta}_i\sim N[\boldsymbol{0},\boldsymbol{\Psi}_\delta],\ \boldsymbol{\xi}_i\sim N[\boldsymbol{0},\boldsymbol{\Phi}], \tag{2}$$

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

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 $\boldsymbol{\Pi}_0=\boldsymbol{I}-\boldsymbol{\Pi}$, which is assumed to be nonsingular, then model M can be written as

$$\boldsymbol{y}_i=\boldsymbol{u}+\boldsymbol{\Lambda}\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i)+\boldsymbol{\delta}_i)+\boldsymbol{\Lambda}\boldsymbol{\xi}_i+\boldsymbol{\varepsilon}_i. \tag{3}$$

To deal with ordered categorical data, suppose $\boldsymbol{y}_i=(\boldsymbol{y}_{o,i},\boldsymbol{y}_{u,i})$, where $\boldsymbol{y}_{o,i}$ ($r\times 1$) and $\boldsymbol{y}_{u,i}$ ($s\times 1$) are vectors corresponding to the observed and unobserved continuous variables, separately. The information of $\boldsymbol{y}_{u,i}$ is given by the observed categorical variables in $\boldsymbol{z}_i$, and the relationship between them is given as follows:

$$z_i = \begin{bmatrix} z_{i1} \\ \vdots \\ z_{is} \end{bmatrix} \text{ if } \begin{cases} \alpha_{1,z_{i1}} < y_{u,i1} \le \alpha_{1,z_{i1}+1} \\ \vdots \\ \alpha_{1,z_{is}} < y_{u,is} \le \alpha_{1,z_{is}+1} \end{cases}, i = 1,\cdots n; k = 1,\cdots s, \tag{4}$$

where $z_{ik} \in \{0,1,\cdots,b_k\}$, and $\boldsymbol{\alpha}_k = \{\alpha_{k,0}, \alpha_{k,1}, \cdots, \alpha_{k,b_k+1}\}$ is a vector of 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 $\boldsymbol{\alpha}_k$. We denote the model defined by (1) to (4) as M. As discussed in Lee and Song [3], without imposing any identification conditions, models with ordered categorical variables are unidentified. So, we fix $\alpha_{k,1}$ and α_{k,b_k} at some pre-assigned 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 pre-assigned values. In the following part of this paper, some notations are used. Let $\boldsymbol{Y}^{obs} = (\boldsymbol{Y}_o^{obs}, \boldsymbol{Y}_u^{obs})$ be the matrix of the observations, where $\boldsymbol{Y}_o^{obs} = (\boldsymbol{y}_{o,1}^{obs}, \cdots, \boldsymbol{y}_{o,n}^{obs})$ and $\boldsymbol{Y}_u^{obs} = (\boldsymbol{y}_{u,1}, \cdots, \boldsymbol{y}_{u,n})$ are matrices of observations corresponding to the observed and unobserved continuous variables, respectively, and $\boldsymbol{y}_i = (\boldsymbol{y}_{o,i}^{obsT}, \boldsymbol{y}_u^T)^T$. 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. Moreover, 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 $\boldsymbol{Z}^{obs}$. Let $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ be the matrix of latent variables, and $\boldsymbol{\Omega}_1 = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_n)$ 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 let $\boldsymbol{\theta}$ be the vector that contains all the unknown parameters involved in (1) to (3), and $\boldsymbol{\alpha}$ be the vector that contains all the thresholds in (4).

B. Lv measure for nonlinear structural equation models with ordered categorical data

Suppose $\boldsymbol{Y}^{obs} = \{y_{ij}^{obs}, i = 1,\cdots,n, j = 1,\cdots,p\}$ is the matrix of observations which come from the exponential family, then the Lv measure [2] is defined by

$$L_v(\boldsymbol{Y}^{obs}) = \sum_{i=1}^{n}\sum_{j=1}^{p}\left[Var(y_{ij}^{rep} \mid \boldsymbol{Y}^{obs}) + v(y_{ij}^{obs} - \mu_{ij})\right] \tag{5}$$

where y_{ij}^{rep} is the replication of y_{ij}^{obs}, $\mu_{ij} = E(y_{ij}^{rep} \mid \boldsymbol{Y}^{obs})$ is the conditional expectation which is taken with respect to the posterior predictive distribution:

$p(y_{ij}^{rep} \mid \boldsymbol{Y}^{obs}) = \int p(y_{ij}^{rep} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$.

However, in the proposed model M, the observations are $(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$, which include ordered categorical data. The distribution of order categorical variable are not belongs to exponential family. Thus, (5) cannot be used directly for this kind of data. In this paper, the method proposed by Chen et al. [4] is used to transform these ordered categorical data into a binary data. Specifically, a new vector $\boldsymbol{z}_{ik}^* = (z_{ik,1}, \cdots, z_{ik,b_k+1})^T$ is defined as follows:

$$z_{ik,j} = \begin{cases} 1, & if \quad z_{ik} = j-1 \\ 0, & otherwise \end{cases}, \quad j = 1,\cdots,b_k+1, k = 1,\cdots,s. \tag{6}$$

From (4), we have

$p(z_{ik,j} = 1) = p(z_{ik} = j-1) = p(\alpha_{k,j-1} < y_{u,ik} \le \alpha_{k,j}) \triangleq p_{ik,j}$, and $p(z_{ik,j} = 0) = 1 - p(z_{ik,j} = 1) = 1 - p_{ik,j}$. It can be shown that $z_{ik,j} \sim Bernoulli(p_{ik,j})$ which belongs to exponential family, and $E(z_{ik,j}) = p_{ik,j}$, $Var(z_{ik,j}) = p_{ik,j}(1 - p_{ik,j})$. According to (6), z_{ij}^{obs} and z_{ij}^{rep} can be transformed to binary vectors $\boldsymbol{z}_{ik}^{obs*}$ and $\boldsymbol{z}_{ik}^{rep*}$, respectively. Then the quadratic loss Lv measure for the ordered categorical data is defined by

$$\begin{aligned} &L_v^q(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) \\ &= \sum_{k=1}^{s}\sum_{i=1}^{n}\left[tr\left\{Var(\boldsymbol{z}_{ik}^{rep*} \mid \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)\right\} + v(\boldsymbol{z}_{ik}^{obs*} - \boldsymbol{\mu}_{ik}^*)^T(\boldsymbol{z}_{ik}^{obs*} - \boldsymbol{\mu}_{ik}^*)\right], \end{aligned} \tag{7}$$

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

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

where $\boldsymbol{e}_j$ is a $(b_k+1)\times 1$ vector with 1 at the j th element and 0 at the others, and

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

It can be shown that (7) can be rewritten as:

$$\begin{aligned} &L_v^q(\boldsymbol{Y}^{obs}, \boldsymbol{Z}^{obs}, M) \\ &= \sum_{k=1}^{s}\sum_{i=1}^{n}\sum_{j=1}^{b_k+1} Var(z_{ik,j}^{rep*} \mid \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) + v(z_{ik,j}^{rep*} - \mu_{ik,j}^*)^2 \end{aligned} \tag{8}$$

where $Var(z_{ik,j}^{rep*} \mid \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)$ is the jth diagonal element of the conditional variance matrix $Var(\boldsymbol{z}_{ik}^{rep*} \mid \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)$ and $\mu_{ik,j}^*$ is the jth element of the conditional expectation $\boldsymbol{\mu}_{ik}^*$. Thus,

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

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 $\boldsymbol{\Lambda}_{u,\eta}(\boldsymbol{\Pi}_0^{-1}\boldsymbol{\Psi}_\delta(\boldsymbol{\Pi}_0^{-1})^T)\boldsymbol{\Lambda}_{u,\eta}{}^T + \boldsymbol{\Psi}_{\varepsilon u}$, where $\boldsymbol{u}_u$, $\boldsymbol{\Lambda}_{u,\eta}$, $\boldsymbol{\Lambda}_{u,\xi}$ and $\boldsymbol{\Psi}_{\varepsilon u}$ are the submatrices of $\boldsymbol{u}$, $\boldsymbol{\Lambda}_\eta$, $\boldsymbol{\Lambda}_\xi$ and $\boldsymbol{\Psi}_\varepsilon$ corresponding to $\boldsymbol{y}_{u,i}$, respectively. Therefore,

$$p(\alpha_{k,j-1} < y_{u,ik}^{rep} \le \alpha_{k,j} \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i, \boldsymbol{y}_{u,i}, M) = \Phi(A_0) - \Phi(A_1)$$

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; $\psi_{\varepsilon uk}$ is the kth

diagonal element of $\boldsymbol{\Psi}_{\varepsilon u}$; $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Similarly, the conditional variance required in (7) is given by

$$Var(z_{ik,j}^{rep*} \mid \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) = \mu_{ik,j}^{*}(1-\mu_{ik,j}^{*}).$$

For the observations $\boldsymbol{Y}_o^{obs}$ corresponding to the observed continuous variables, the Lv measure defined by (5) is used. Therefore, given observations $(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs})$, the Lv measure for the model M can be defined by

$$\begin{aligned} &L_v(\boldsymbol{Y}^{obs}, \boldsymbol{Z}^{obs}, M) \\ &= \sum_{k=1}^{r}\sum_{i=1}^{n}\left[Var(y_{ik}^{rep} \mid \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) + v(y_{ik}^{rep} - \mu_{ik})^2 \right] \\ &+ \sum_{k=1}^{s}\sum_{i=1}^{n}\sum_{j=1}^{b_k+1}\left[Var(z_{ik,j}^{rep*} \mid \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M) + v(z_{ik,j}^{rep*} - \mu_{ik,j}^{*})^2 \right], \end{aligned} \quad (9)$$

where $\mu_{ik} = E(y_{ik}^{rep} \mid \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M)$. Let $\boldsymbol{u}_o$, $\boldsymbol{\Lambda}_{o,\eta}$, $\boldsymbol{\Lambda}_{o,\xi}$ and $\boldsymbol{\Psi}_{\varepsilon u}$ are submatrices of $\boldsymbol{u}$, $\boldsymbol{\Lambda}_\eta$, $\boldsymbol{\Lambda}_\xi$ and $\boldsymbol{\Psi}_\varepsilon$ corresponding to $\boldsymbol{y}_{o,i}^{obs}$, respectively. Then we get

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

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 (9) is given by

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

Due to the existence of intractable integrals in calculating the conditional expectation and variance, we cannot get a closed form of the Lv measure. Therefore, Markov Chain Monte Carlo (MCMC) methods are used to calculate the Lv measure for the proposed SEM.

C. Computation of the Lv Measure

From its definition, the Lv measure can be estimated with a sufficiently large number of random observations $\{\boldsymbol{\theta}^{(g)}, \boldsymbol{\Omega}^{(g)}, \boldsymbol{\alpha}^{(g)}, \boldsymbol{Y}_u^{(g)}; g = 1, \cdots, G\}$, which are generated from the conditional distribution $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{Y}_u \mid \boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, \boldsymbol{M})$. To generate this sample, Gibbs sampler algorithm [5] given below is used. With current values $(\boldsymbol{\theta}^{(g)}, \boldsymbol{\Omega}^{(g)}, \boldsymbol{\alpha}^{(g)}, \boldsymbol{Y}_u^{(g)})$:

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

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

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

After obtaining convergence, the random sample of latent variables and unknown parameters can be collected. For $i = 1, \cdots, n$ and $k = 1, \cdots, r$, let

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

$$\hat{\mu}_{ik} = \frac{1}{G}\sum_{g=1}^{G}\mu_{ik}^{(g)} \text{ , and } \hat{\sigma}_{ik} = \frac{1}{G}\sum_{g=1}^{G}[\psi_{\varepsilon ok}^{(g)} + (\mu_{ik}^{(g)})^2] - \hat{\mu}_{ik}^{\,2}.$$

Moreover, for $k = 1, \cdots, s$ and $j = 1, \cdots, b_k + 1$, let

$$A_{ik,j}^{(g)} = \frac{\alpha_{k,j}^{(g)} - \{u_{u,k}^{(g)} + \boldsymbol{\Lambda}_{u,\eta k}^{(g)}(\boldsymbol{\Pi}_0^{(g)})^{-1}\boldsymbol{\Gamma}^{(g)}\boldsymbol{F}(\boldsymbol{\xi}_i^{(g)}) + \boldsymbol{\Lambda}_{u,\xi k}^{(g)}\boldsymbol{\xi}_i^{(g)}\}}{\psi_{\varepsilon uk}^{(g)}},$$

$$\mu_{ik,j}^{(g)*} = \Phi(A_{ik,j}^{(g)}) - \Phi(A_{ik,j-1}^{(g)}) \text{ , } \hat{\mu}_{ik,j} = \sum_{g=1}^{G}\mu_{ik,j}^{(g)*}/G \text{ , } \hat{\sigma}_{ik,j}^{*} = \hat{\mu}_{ik,j}(1-\hat{\mu}_{ik,j}).$$

Then the estimate of the Lv measure for the proposed model can be given by

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

To obtain the conditional distributions in the Gibbs sampler, the prior distributions of the unknown parameters are needed. In this paper, the following commonly used conjugate type prior distributions are used:

$$p(\boldsymbol{u}) \triangleq N[\boldsymbol{u}_0, \boldsymbol{\Sigma}_0], \qquad p(\boldsymbol{\Phi}) \triangleq IW_{q_2}[\boldsymbol{R}_0, \rho_0], \quad (10)$$

$$\begin{aligned} &p(\psi_{\delta k}^{-1}) \triangleq \text{Gamma}[\alpha_{0\delta k}, \beta_{0\delta k}], \\ &p(\boldsymbol{\Lambda}_{\omega k} \mid \psi_{\delta k}) \triangleq N[\boldsymbol{\Lambda}_{0\omega k}, \psi_{\delta k}\boldsymbol{H}_{0\omega k}], k = 1, \cdots, q_1, \end{aligned} \quad (11)$$

$$p(\psi_{\varepsilon j}^{-1}) \triangleq \text{Gamma}[\alpha_{0\varepsilon j}, \beta_{0\varepsilon j}], \quad p(\boldsymbol{\Lambda}_j \mid \psi_{\varepsilon j}) \triangleq N[\boldsymbol{\Lambda}_{0j}, \psi_{\varepsilon j}\boldsymbol{H}_{0j}], j = 1, \cdots, p. \quad (12)$$

where the parameters in the above prior distributions are hyperparameters whose values are assumed to be given. For $k \neq h$, it is assumed that $(\psi_{\varepsilon k}, \boldsymbol{\Lambda}_k)$ and $(\psi_{\varepsilon h}, \boldsymbol{\Lambda}_h)$ are independent; for $j \neq h$, $(\psi_{\delta j}, \boldsymbol{\Lambda}_{\omega j})$ and $(\psi_{\delta h}, \boldsymbol{\Lambda}_{\omega h})$ are assumed to be independent. For the unknown thresholds in $\boldsymbol{\alpha}$, the following non-informative prior distribution is used:

$$p(\boldsymbol{\alpha}_k) = p(\alpha_{k,1}, \cdots, \alpha_{k,b_k+1}) \propto C, \quad \alpha_{k,2} < \cdots < \alpha_{k,b_k+1}, k = 1, \cdots, s.$$

where C is a constant. With these prior distributions, the posterior distributions required in the Gibbs sampler can be obtained. However, the conditional distributions required in **step a** and **step c** are non-standard and complex, the Metropolis-Hastings (MH) algorithm [6][7] is used. To save space, the conditional distributions are not presented, details can be found in [1].

D. Calibration Distribution

As pointed out by [2], 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 selection 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 $L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)$ denote the Lv measure of the candidate model M_c, and $L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_t)$ denote the Lv measure of the true model M_t. Then given v, the difference of the Lv measures between the candidate model M_c and the true model M_t is defined as

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

Then the calibration distribution is defined as the marginal distribution of $D_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)$. From the definition, given v, $D_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)$ is a random variable of $\boldsymbol{Y}^{obs}$. Thus, the marginal distribution is computed with respect to the prior

predictive distribution of $\boldsymbol{Y}^{obs}$ under the true model M_t:

$$p_t(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}) = \int p(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs} \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, M_t) p(\boldsymbol{\theta}, \boldsymbol{\alpha} \mid M_t) d\boldsymbol{\theta} d\boldsymbol{\alpha}$$

where $p(\boldsymbol{\theta}, \boldsymbol{\alpha} \mid M_t)$ denotes the prior distribution of the unknown parameters under the true model M_t. We denote the calibration distribution by

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

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 the highest probability density (HPD) interval, the mean μ_v, and the standard deviation SD_v of the calibration distribution. Here, HPD interval denotes the shortest credible interval that means the interval with the highest posterior density, and it can be computed by using a Monte Carlo (MC) method [8]. μ_v measures, on the average, how close the candidate model and the true model are. SD_v measures the variability of calibration distribution. [3] 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.

For the proposed model, we cannot get a closed form of the calibration distribution. So MCMC methods are used again to estimate the calibration distributions of the models under consideration. The specific procedure for the estimation of PL_c is given as follows: Generate $(\tilde{\boldsymbol{\Omega}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\alpha}})$ from the prior predictive distribution $p(\boldsymbol{\Omega} \mid \boldsymbol{\theta}, M_t) p(\boldsymbol{\theta} \mid M_t) \quad p(\boldsymbol{\alpha} \mid M_t)$; Generate $\tilde{\boldsymbol{Y}}$ from distribution $p(\boldsymbol{Y} \mid \tilde{\boldsymbol{\Omega}}, \tilde{\boldsymbol{\theta}}, , \tilde{\alpha}, M_t)$; Set $\boldsymbol{Y}^{obs} = \tilde{\boldsymbol{Y}}$, calculate the Lv measures $L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)$ and $L_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_t)$, then calculate $D_v(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c)$. Repeat these three steps H times, we can collect a sample of $\{D_v^{(h)}(\boldsymbol{Y}_o^{obs}, \boldsymbol{Z}^{obs}, M_c), h = 1, \cdots, H\}$. Based on this sample, the calibration distribution PL_c can be estimated via the kernel density estimation method [9], and then the summaries of PL_c can be easily obtained. Since the true model is usually unknown in practical applications, the model with the smallest Lv measure will be considered as the true model M_t.

III. A Simulation Study

The observations $\{\boldsymbol{y}_{o,i}, \boldsymbol{z}_i; i = 1, \cdots, n\} (n = 300)$ are generated from the following model:

$M_0 : \boldsymbol{y}_i = \boldsymbol{u} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i$, 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 $\boldsymbol{y}_{u,i}$ is generated from (4) through threshold $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_4)$. The specification of the loading matrix is

$$\boldsymbol{\Lambda}^T = \begin{pmatrix} 1^* & \lambda_{21} & \lambda_{31} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 1^* & \lambda_{52} & \lambda_{62} & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1^* & \lambda_{83} & \lambda_{93} \end{pmatrix}$$

where the elements with asterisks are fixed. The true values of the unknown parameters in this model are given as: $\lambda_{21} = \lambda_{31} = 0.8$, $\lambda_{52} = \lambda_{62} = 0.7$, $\lambda_{83} = \lambda_{93} = 0.8$, $\gamma_1 = 0.6$, $\gamma_2 = 0.6$, $\gamma_3 = -0.5$ $\boldsymbol{u} = (0, \cdots, 0)^T$, $\phi_{11} = \phi_{22} = 1.0$, $\phi_{12} = \phi_{21} = 0.2$, $\psi_{\varepsilon 1} = \cdots = \psi_{\varepsilon 9} = 0.5$, $\psi_\delta = 0.5$, and $\boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_4 = (-1.0^*, -0.6, 0.6, 1.0^*)^T$, in which the elements with asterisks are fixed. Four SEMs denoted by $M_1, \cdots, M_4$ are considered as competing models. The measurement equations of these models are the same as that of M_0. Their structural equations are

$$M_1 : \eta_{1i} = \gamma_{1,1}\xi_{1,i1} + \gamma_{1,2}\xi_{1,i2} + \gamma_{1,3}\xi_{1,i1}^{\ 2} + \gamma_{1,4}\xi_{1,i2}^{\ 2} + \gamma_{1,5}\xi_{1,i1}\xi_{1,i2} + \delta_{1,i},$$

$$M_2 : \eta_{2i} = \gamma_{2,1}\xi_{2,i1} + \gamma_{2,2}\xi_{2,i2} + \delta_{2,i},$$

$$M_3 : \eta_{3i} = \gamma_{3,1}\xi_{3,i1} + \gamma_{3,2}\xi_{3,i2} + \gamma_{3,3}\xi_{3,i1}^{\ 2} + \delta_{1,i},$$

$$M_4 : \eta_{4i} = \gamma_{4,1}\xi_{4,i1} + \gamma_{4,2}\xi_{4,i2} + \gamma_{4,3}\xi_{4,i2}^{\ 2} + \delta_{1,i}.$$

The prior distributions given in (10) to (12) are used. 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: The means in the normal prior distributions are taken as the true values of the corresponding parameters, and the covariance matrices are taken as the identity matrices with corresponding dimensions; 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_{12} and ϕ_{22}; 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 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 $\beta_{0k} = \beta_{0\delta} = 5$.

For each replication, a total of $R = 2000$ observations are collected after 2000 burn-in iterations. The results are obtained on the basis of 100 replications. The calibration summaries are given in TABLE 1, where $\mu_{0.5}$, $SD_{0.5}$, 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 1, we see that under each type of prior inputs, $\mu_{0.5}$ corresponding to M_2, M_3 and M_4 are substantially larger than zero. In addition, the 95% HPD intervals corresponding to these models do not include zero. Thus, we can conclude that M_2, M_3 and M_4 are far from the true model, and M_0 performs much better than them. However, for model M_1, $\mu_{0.5}$ is not significantly different from zero given the large value of $SD_{0.5}$ 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 M_1 performs similar as M_0, according to the parsimonious principle, the simpler model M_0 is selected under each given type of prior inputs. The estimated calibration distributions presented in Fig. 1 (a) and Fig. 1(b) also agree with these conclusions.

TABLE 1. CALIBRATION SUMMARIES

	Model	$\mu_{0.5}$	$SD_{0.5}$	95%HPD
Prior I	M_1	1.118	10.463	(-14.35,22.74)
	M_2	44.726	21.968	(5.489,79.53)
	M_3	41.155	25.904	(8.724,84.18)
	M_4	46.998	34.075	(2.47,88.519)
Prior II	M_1	1.91	20.953	(-19.854,77.5)
	M_2	44.237	22.139	(4.492,89.98)
	M_3	39.869	21.143	(-0.313,83.85)
	M_4	41.832	22.702	(5.707,100.52)

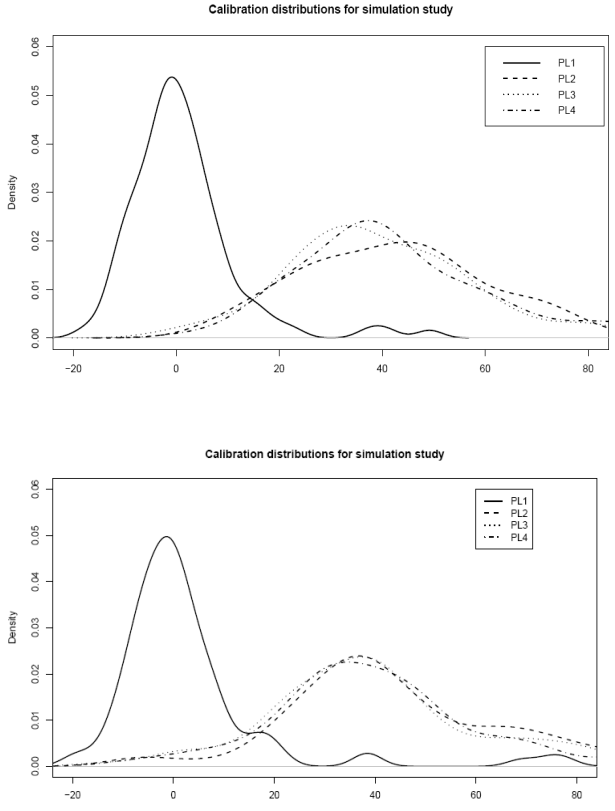

Fig1.Calibration distributions in simulation study: (a) under Prior I, (b)under Prior II

IV. 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 [1]. The aim of this section is to apply the Lv measure to model selection in the analysis of this QOL data. The instrument WHOQOL-100 for measuring QOL given in [10] 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 Lv measure, we only analyze a synthetic data set with sample size n = 338. 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

$$\boldsymbol{y} = \boldsymbol{\Lambda}_1\boldsymbol{\omega}_1 + \boldsymbol{\varepsilon},$$

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

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

The structural equation of M_1 is given by

$$\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \gamma_4\xi_4 + \delta.$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3, \xi_4)^T$ and δ are independently distributed as $N[\boldsymbol{0}, \boldsymbol{\Phi}_1]$ and $N[0, \sigma_{1\delta}^2]$, respectively. The measurement equation of M_2 is defined by

$$\boldsymbol{y} = \boldsymbol{\Lambda}_2\boldsymbol{\omega}_2 + \boldsymbol{\varepsilon},$$

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

$$\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \delta.$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^T$ and δ are independently distributed as $N[\boldsymbol{0}, \boldsymbol{\Phi}_2]$ 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 $\boldsymbol{y}$ and $\boldsymbol{z}$ is defined by equation (4). The threshold are given by $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{26})^T$, where $\boldsymbol{\alpha}_k = (\alpha_{k1}, \cdots, \alpha_{k6})$, $\alpha_{k1} = -\infty$, $\alpha_{k6} = \infty$, $k = 1, \cdots 26$. 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 [1] for more details.

To calculate the Lv measure, the conjugate prior distributions are used. The hyperparameter values corresponding to the prior distributions of the unknown loadings in $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ 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 $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ are $\rho_1 = \rho_2 = 30$. $\boldsymbol{R}_{01}^{-1} = 8\boldsymbol{I}_4$, $\boldsymbol{R}_{02}^{-1} = 8\boldsymbol{I}_3$, respectively; let $\boldsymbol{I}_d$ be an identity matrix with

dimension d, we take $\boldsymbol{H}_{0\omega k1} = 0.25\boldsymbol{I}_4$, and $\boldsymbol{H}_{0\omega k2} = 0.25\boldsymbol{I}_3$; $\alpha_{0\varepsilon k1} = \alpha_{0\varepsilon k2} = \alpha_{0\delta k1} = \alpha_{0\delta k1} = 10$ and $\beta_{0\varepsilon k1} = \beta_{0\varepsilon k2} = \beta_{0\delta k1} = \beta_{0\delta k1} = 10$. In the Gibbs sampling in computing the Lv measure and the estimation of unknown parameters, we take J=2000 observations after a burn-in phase of 4000 iterations. Lv measure is 7273.01 for M_1 and 7343.826 for M_2. As the value of the Lv 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 2, and the density of the calibration distribution is given in Fig. 3.

TABLE 2. CALIBRATION SUMMARIES FOR REAL EXAMLE

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

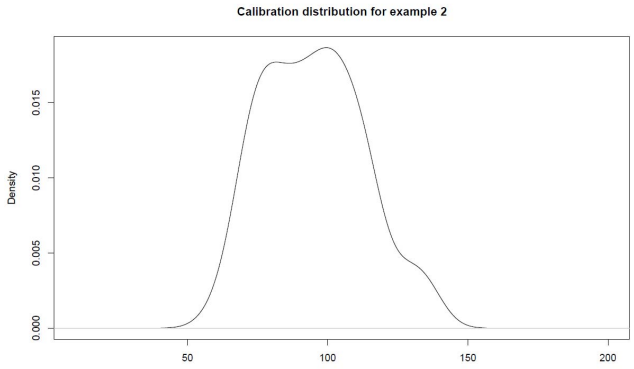

Fig. 2 Calibration Distribution for Real Example

We see that the mean of the difference between the Lv measures of M_1 and M_2 is larger than zero, and the 95\% HPD interval does not include zero. Therefore, M_1 is selected. The estimation of the the unknown parameters are given in Table 3.

To compare M_1 and M_2 by using Bayes factor, path sampling [11] is applied. First, we will compare model M_1 with the following model M_0:

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

where $\boldsymbol{\varepsilon} \sim N[\boldsymbol{0}, \boldsymbol{\Psi}_\varepsilon]$ and $\boldsymbol{\Psi}_\varepsilon$ is a diagonal matrix. We obtain logB10 =81:36. Similarly, M_2 and M_0 can be compared via the path sampling procedure, and logB20 = 57:85; which means that M_1 and M_2 are both better than M_0. Furthermore, from the above result, logB12 is equal to 23.51. Therefore, M_1 is selected.

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, we can get the same conclusion as given by Lv measure.

V. DISCUSSION

From the numerical studies given in the previous section, the Lv 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 Lv 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 Lv measure is quite simple and fast. Moreover, besides considering the model with the smallest value of the Lv measure, the corresponding calibration distribution is also used to help making decision. Therefore, the Lv measure provides better alternative method for model selection of SEMs.

TABLE 3. BAYESIAN ESTIMATES OF UNKNOWN PARAMETERS IN M1

Parameter	EST	SD	Parameter	EST	SD	Parameter	EST	SD
λ_{21}	0.85	0.07	ψ_δ	0.25	0.03	$\psi_{\varepsilon 17}$	0.96	0.09
λ_{42}	0.91	0.09	γ_1	0.76	0.09	$\psi_{\varepsilon 18}$	0.52	0.06
λ_{52}	1.06	0.08	γ_2	0.37	0.1	$\psi_{\varepsilon 19}$	0.53	0.06
λ_{62}	1.14	0.09	γ_3	0.14	0.11	$\psi_{\varepsilon 20}$	0.67	0.07
λ_{72}	0.79	0.09	γ_4	-0.03	0.11	$\psi_{\varepsilon 21}$	0.7	0.07
λ_{82}	1.26	0.08	$\psi_{\varepsilon 1}$	0.39	0.05	$\psi_{\varepsilon 22}$	0.7	0.07
λ_{92}	1.14	0.08	$\psi_{\varepsilon 2}$	0.42	0.05	$\psi_{\varepsilon 23}$	0.74	0.07
$\lambda_{11,3}$	0.8	0.09	$\psi_{\varepsilon 3}$	0.62	0.07	$\psi_{\varepsilon 24}$	0.57	0.06
$\lambda_{12,3}$	0.72	0.08	$\psi_{\varepsilon 4}$	0.61	0.07	$\psi_{\varepsilon 25}$	0.71	0.07
$\lambda_{13,3}$	0.75	0.09	$\psi_{\varepsilon 5}$	0.46	0.05	$\psi_{\varepsilon 26}$	0.66	0.07
$\lambda_{14,3}$	1	0.08	$\psi_{\varepsilon 6}$	0.4	0.05	ϕ_{11}	0.49	0.06
$\lambda_{15,3}$	0.86	0.08	$\psi_{\varepsilon 7}$	0.7	0.06	ϕ_{12}	0.35	0.04
$\lambda_{17,4}$	0.28	0.09	$\psi_{\varepsilon 8}$	0.28	0.03	ϕ_{13}	0.22	0.04
$\lambda_{18,4}$	0.95	0.01	$\psi_{\varepsilon 9}$	0.39	0.04	ϕ_{14}	0.31	0.04
$\lambda_{20,5}$	0.8	0.08	$\psi_{\varepsilon 10}$	0.47	0.05	ϕ_{22}	0.58	0.07
$\lambda_{21,5}$	0.77	0.09	$\psi_{\varepsilon 11}$	0.65	0.07	ϕ_{23}	0.38	0.05
$\lambda_{22,5}$	0.76	0.09	$\psi_{\varepsilon 12}$	0.71	0.07	ϕ_{24}	0.39	0.05
$\lambda_{23,5}$	0.71	0.09	$\psi_{\varepsilon 13}$	0.7	0.07	ϕ_{33}	0.59	0.08
$\lambda_{24,5}$	0.97	0.1	$\psi_{\varepsilon 14}$	0.45	0.05	ϕ_{34}	0.38	0.05
$\lambda_{25,5}$	0.77	0.09	$\psi_{\varepsilon 15}$	0.57	0.06	ϕ_{44}	0.54	0.07
$\lambda_{26,5}$	0.84	0.1	$\psi_{\varepsilon 16}$	0.46	0.06			

REFERENCES

[1] S.Y. Lee, "Structural Equation Modeling: A Bayesian Approach," *Wiley, 2007.*

[2] J.G. Ibrahim, M.H. Chen and D. Sinha, "Criterion based methods for Bayesian model assessment," *Statistica Sinica,* vol. 11, pp. 419-443, 2001.

[3] S. Y. Lee and X.Y. Song,"Bayesian model selection of nonlinear latent variable models with missing continuous and ordinal categorical data," *British Journal of Mathematical and Statistical Psychology,*vol 57, pp.131-150,2004.

[4] M. H. Chen, D.K. Dey, and J.G. Ibrahim, J. G., "Bayesian criterion based model assessment for categorical data," *Biometrika*, vol 91, pp.45-63,2004.

[5] S. Geman, and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, 1984.

[6] N. Metropolis, A.W. Rosenbluth , M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equations of state calculations by fast computing machine. *Journal of Chemical Physics,*" vol. 21, pp.1087-1091, 1953.

[7] M. W.K. Hastings, "Monte Carlo sampling methods using Markov chains and their application," *Biometrika,* vol. 57, pp. 97-109, 1970.

[8] M.H. Chen, and Q.M. Shao, "Monte Carlo estimation of Bayesian credible and HPD intervals," *Journal of Computational and Graphical Statistics,* vol. 8, pp. 69-92, 1999.

[9] B.W. Silverman, "Density Estimation for statistics and data analysis,". *London:Chapman and Hall, 1986.*

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

[11] A. Gelman and X.L. Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. Statistical Science, 1998, vol 13, 163-185.

Y.X. Li was born in Yunnan, China, in May 1981. She received her PhD degree in statistics from the Chinese University of Hong Kong (CUHK) in 2010. Her major field of study is applied statistics. From July 2006 to July 2007 Dr Li worked at the School of Finance, Yunnan University of Finance and Economics, as a Teaching Assistant. After that she moved to Hong Kong to pursue her PhD in CUHK. In the following three years, Dr Li worked as a Teaching Assistant at the Department of Statistics in CUHK. After completed the PhD study, she returned to Yunnan University of Finance and Economics, and was appointed as an Instructor at the Department of Insurance, School of Finance. Her current research interests include model selection, actuarial science, and data analysis.

Dr. Li is a member of the International Association of Computer Science and Information Technology (IACSIT).