Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations

Xin-Yuan Song and Sik-Yum Lee*

Department of Statistics, Chinese University of Hong Kong

The main purpose of this paper is to develop a Bayesian approach for the multi-sample factor analysis model with continuous and polytomous variables. Joint Bayesian estimates of the thresholds, the factor scores and the structural parameters subjected to some simple constraints across groups are obtained simultaneously. The Gibbs sampler is used to produce the joint Bayesian estimates. It is shown that the conditional distributions involved in the implementation are the familiar uniform, gamma, normal, univariate truncated normal and Wishart distributions. The Bayes factor is introduced to test hypotheses involving constraints among the structural parameters of the factor analysis models across groups. Two procedures for computing the test statistics are developed, one based on the Schwarz criterion (or Bayesian information criterion), while the other computes the posterior densities and likelihood ratios by means of draws from the appropriate conditional distributions via the Gibbs sampler. The empirical performance of the proposed Bayesian procedure and its sensitivity to prior distributions are illustrated by some simulation results and two real-life examples.

1. Introduction

Factor analysis is an important technique in behavioural science research for assessing the interdependence, causations and correlations among observed variables and latent factors. Traditionally, most analyses of the model have been carried out within the framework of structural equation modelling, under the assumption that the observed variables are continuous. However, in practical applications, owing to the nature of the variables or the design of questionnaires, many variables are polytomous. The likelihood function which takes into account this polytomous nature involves some complicated multiple integrals. Hence, direct computation of the maximum likelihood estimate is tedious (see Lee, Poon & Bentler, 1990). LISCOMP (Muthén, 1987) and Lee, Poon & Bentler (1995) used some multi-stage approaches to reduce the computational burden of the multiple integrals. In practice, since the size of the weight matrix involved in the multi-stage estimation increases rapidly with the dimension of the model, numerical difficulties are encountered when dealing with high-dimensional models. To avoid the heavy computations involved in evaluating

* Requests for reprints should be addressed to Professor S. Y. Lee, Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong.
the multiple integrals, a Monte Carlo EM algorithm was proposed by Shi & Lee (2000) for obtaining the maximum likelihood estimate, while Shi & Lee (1998) developed a Bayesian approach with the Gibbs sampler and demonstrated some advantages of the Bayesian procedure over the multi-stage methods.

For the factor analysis model with continuous data, an important generalization is to extend the model to permit analysis of multiple groups or populations of individuals simultaneously. Multi-sample analysis is important in various applications, such as cross-cultural research. Almost all currently available software in the field includes options for carrying out multi-sample analysis; see, for example, EQS (Bentler, 1992), and LISREL (Jöreskog & Sörbom, 1996). However, as far as we know, very limited theoretical or computational results have been achieved for the multi-sample factor analysis model with polytomous data. Since the number of thresholds and structural parameters involved in multi-sample models is larger than for single-sample models, the computational difficulty caused by the huge weight matrix is more severe in multi-stage methods. In this paper, we investigate a multi-sample factor analysis model with continuous and polytomous variables. The underlying structural parameters are allowed to satisfy some simple equality constraints. Hence, the analysis is capable of testing any degree of invariance in the model—from one extreme where nothing is invariant, to the other extreme where everything is invariant. Our first main objective is to obtain constrained estimates of the unknown structural parameters, thresholds and factor scores. To deal with the computational difficulties caused by the nature of the polytomous variables, a Bayesian approach is developed with the Gibbs sampler (Geman & Geman, 1984) algorithm, in which the continuous latent measurements and the latent factor scores in different groups are treated as hypothetical missing data. Non-informative priors are used for the thresholds and conjugate priors are used for the structural parameters. The present approach is thus a generalization of the Bayesian method given in Shi & Lee (1998).

Beyond the estimation problem, an important topic in multi-sample analysis is to test various relationships across the different groups. For example, in cross-cultural research, it is very interesting to investigate whether the measurement items (which are often on Likert scales of a polytomous nature) of different cultures are distinct by testing the invariance of the factors’ loading matrices and covariance matrices across groups (see, for example, Byrne, Shavelson & Muthén, 1989; Fava & Velicer, 1992; Millsap, 1997). To achieve this goal requires statistical methods for testing various hypotheses in the multi-sample factor analysis model with polytomous variables. But very little has so far been done on this important topic. Note that, as stressed by Meng (1994), the posterior predictive $p$-value (Rubin, 1984; Gelman, Meng & Stern, 1996) in the general Bayesian analysis context serves only as a measure of the discrepancy between the posited model and the observed data. It provides information to enable the goodness-of-fit to be assessed for a single model; but may not be suitable for comparing different models (see also Carlin & Louis, 1996).

The second main objective of this paper is to introduce the Bayes factor (see Berger, 1985) to test various hypotheses about the multi-sample factor analysis model with continuous and polytomous variables. The computation of the Bayes factor involves multiple integrals which are the marginal densities of the data under different hypotheses. In simple cases, these integrals may be evaluated analytically. More often, they are intractable and have to be computed by numerical methods. Hence, the computation of the Bayes factor is non-trivial and it is still the focus of much attention in recent Bayesian literature; see, for example, Aitkin (1991), Raftery (1993, 1995), Newton & Raftery (1994), Chib (1995), Draper (1995),
O’Hagan (1995), Kass & Raftery (1995), Kass & Wasserman (1995, 1996), and DiCiccio, Kass, Raftery & Wasserman (1997). In the context of a single-group structural equation model with continuous variables, Raftery (1993) pointed out the difficulties of using the common significance tests based on \( p \) values, and proposed the Bayes factor for hypothesis testing and model selection. He adapted the rough but simple Bayesian information criterion (BIC) approximation, which depends only on the final likelihood function value provided by standard software such as LISREL (Jöreskog & Sörbom, 1996). For the current Bayesian analysis with the multi-sample model that involves mixed continuous and polytomous variables, the computation of the Bayes factor is much more difficult. Even the simple Schwarz criterion (or BIC) is not in closed form.

In this paper, we develop a computational procedure that utilizes the draws from the Gibbs sampler to compute the Schwarz criterion (or BIC). Moreover, inspired by Chib’s (1995) procedure for computing the marginal density, another procedure is developed. In this procedure, the Gibbs sampler is used to compute the posterior densities and the likelihood ratios that are required in the computation of the Bayes factor. Since thresholds are not involved in the hypotheses of interest, the prior distributions of these nuisance parameters do not greatly affect the results (see, for example, Kass & Vaidyanathan, 1992; Kass & Raftery, 1995). Sensitivity to the choices of hyperparameter values in the conjugate prior distributions of the structural parameters in the hypotheses is studied.

The paper is organized as follows. The basic multi-sample factor analysis model with mixed continuous and polytomous variables is described in Section 2. In Section 3, an algorithm based on the Gibbs sampler (Geman & Geman, 1984) is developed to produce joint Bayesian estimates of the thresholds, factor scores and structural parameters that satisfy some simple constraints across groups. The Bayes factor and its computation for hypothesis testing are discussed in Section 4. Two procedures, one based on the Schwarz criterion (or BIC) and the other based on the computation of the marginal densities, are developed here. To illustrate the proposed methodologies and to study the sensitivity of the priors relating to hypothesis testing, results from some simulation results and two real examples are reported in Section 5. A discussion is given in Section 6. Some technical details are given in the Appendices.

2. Model description

Consider a set of \( G \) populations which may be different nations, states or regions, cultural or socio-economic groups, groups receiving different treatments, etc. The definition of the multi-sample factor analysis model is given by

\[
\begin{bmatrix}
    x_i^{(g)} \\
    y_i^{(g)}
\end{bmatrix} = \begin{bmatrix}
    \Lambda_1^{(g)} \\
    \Lambda_2^{(g)}
\end{bmatrix} \xi_i^{(g)} + \epsilon_i^{(g)}, \quad i = 1, \ldots, n_g, \quad g = 1, \ldots, G,
\]

where \( \Lambda_1^{(g)}(r \times q) \) and \( \Lambda_2^{(g)}(s \times q) \) are factor loading matrices, \( \xi_i^{(g)} \) is a \( q \times 1 \) vector of latent factor scores with distribution \( N(0, \Phi^{(g)}) \), and \( \epsilon_i^{(g)} \) is a \( p \times 1 \) vector of error measurements with distribution \( N(0, \Psi^{(g)}) \), with \( \Psi^{(g)} = \text{diag}(\psi_{11}^{(g)}, \ldots, \psi_{pp}^{(g)}) \); and \( \epsilon_i^{(g)} \) is independent of \( \xi_i^{(g)} \). Let \( u_i^{(g)} \). Let \( u_i^{(g)} = (x_i^{(g)}, y_i^{(g)})' \) and \( \Lambda^{(g)} = (\Lambda_1^{(g)}, \Lambda_2^{(g)})' \); then the covariance structure of \( u_i^{(g)} \) is given by \( \Sigma^{(g)} = \Sigma(\theta^{(g)}) = \Lambda^{(g)} \Phi^{(g)} \Lambda^{(g)}' + \Psi^{(g)} \), for \( g = 1, \ldots, G \). It is assumed that \( u_i^{(g)}, i = 1, \ldots, n_g, \) are independent and that observations from different groups are independent.
Suppose $y^{(g)}$ is unobservable, and its information is given by the observable polytomous random vector $z^{(g)}$. The relationship between $y^{(g)}$ and $z^{(g)}$ is given by

$$z^{(g)} = \begin{bmatrix} z_1^{(g)} \\ \vdots \\ z_s^{(g)} \end{bmatrix}$$

if $\alpha_{k,z_i}^{(g)} < y_1^{(g)} \leq \alpha_{k,z_i+1}^{(g)}, \ldots, \alpha_{s,z_s}^{(g)} < y_s^{(g)} \leq \alpha_{s,z_s+1}^{(g)},$ \hspace{1cm} (2)

where $z_{k}^{(g)}$ is an integral value in the set $\{0, 1, \ldots, b_k\}$ for $k = 1, \ldots, s$; and $\alpha_{k,0}^{(g)} = -\infty$, $\alpha_{k,b_k+1}^{(g)} = \infty$. Hence, for the $k$th variable, there are $b_k + 1$ categories and the $\alpha_{k,j}^{(g)}$ are the unknown thresholds that define the categories. Naturally, we assume the numbers of continuous variables, polytomous variables, and thresholds are the same across groups. In this model, there are two types of unknown parameters: the thresholds, and the structural parameters in the factor analytic models. For most applications it is the structural parameters that are mainly of interest, not the nuisance threshold parameters. For instance, in testing the invariance of measurement items for different groups, the null hypothesis involves the loading matrices and covariance matrices of the latent factors, while the thresholds are not involved.

It has been pointed out by Lee et al. (1990) that single-sample models with polytomous variables are not identified without imposing identification conditions. This is also the case for multi-sample models. To solve this problem, we use the common method (see, for example, Lee et al., 1995; Shi & Lee, 1998) of fixing some thresholds at pre-assigned values. Moreover, to identify the covariance structures, we follow the common practice in confirmatory factor analysis of fixing appropriate elements of $\Lambda^{(g)}$ at pre-assigned values. For convenience, we assume that the positions of the fixed elements are the same for each group.

3. Bayesian estimation of the multi-sample factor analysis model

In this section, a Bayesian procedure for estimating the thresholds and the constrained structural parameters will be discussed. Direct estimates of the latent factor scores are also produced as a by-product. For the $g$th group, let $X^{(g)} = (x_1^{(g)}, \ldots, x_n^{(g)})$ and $Z^{(g)} = (z_1^{(g)}, \ldots, z_{n_g}^{(g)})$ be the observed continuous and polytomous data matrices, $Y^{(g)} = (y_1^{(g)}, \ldots, y_{n_g}^{(g)})$ and $F^{(g)} = (\xi_1^{(g)}, \ldots, \xi_{n_g}^{(g)})$ be matrices of the latent continuous data and the latent factor scores, $\alpha_k^{(g)}$ be the vector which contains all the unknown thresholds corresponding to the polytomous variable $z_{k}^{(g)}$, and $\alpha^{(g)} = (\alpha_1^{(g)}, \ldots, \alpha_s^{(g)})'$. Moreover, let $X = (X^{(1)}, \ldots, X^{(G)}), \quad Z = (Z^{(1)}, \ldots, Z^{(G)}), \quad Y = (Y^{(1)}, \ldots, Y^{(G)}), \quad F = (F^{(1)}, \ldots, F^{(G)})$ and $n = n_1 + \cdots + n_G$. In connection with the null hypotheses that involve comparisons of the multi-sample covariance structures, it is often interesting to impose the following simple constraints on the structural parameters (see, for example, Jöreskog & Sörbom, 1996):

$$\Lambda^{(1)} = \ldots = \Lambda^{(G)}; \quad (3)$$

$$\Lambda^{(1)} = \ldots = \Lambda^{(G)}, \quad \Phi^{(1)} = \ldots = \Phi^{(G)}; \quad (4)$$

$$\Lambda^{(1)} = \ldots = \Lambda^{(G)}, \quad \Phi^{(1)} = \ldots = \Phi^{(G)}, \quad \Psi^{(1)} = \ldots = \Psi^{(G)}. \quad (5)$$

For models associated with the different constraints specified in (3), (4) or (5), the parameters
The effect of taking this improper prior in computing the Bayes factor for hypothesis testing

First, consider the prior distribution of the less interesting nuisance parameters $\alpha$. For $g \neq h$, it is natural to assume that the prior distributions of $\alpha^{(g)}$ and $\alpha^{(h)}$ are independent. The following non-informative prior distribution is used:

$$p(\alpha^{(g)}) \propto \text{constant}, \quad g = 1, \ldots, G.$$  

The effect of taking this improper prior in computing the Bayes factor for hypothesis testing

Bayesian estimation and test for factor analysis model in several populations

241
will be discussed in more detail in the next section. Now, consider the conditional distribution \([\alpha|X,Z,\theta,Y,F]\). Since the data from different groups are independent, we need only find the marginal conditional distributions for \(\alpha^{(g)}, \ g = 1, \ldots, G\), and \([\alpha^{(g)}|X,Z,\theta,Y,F] = [\alpha^{(g)}|X^{(g)},Z^{(g)},\theta^{(g)},Y^{(g)},F^{(g)}]\). Since \(\Psi^{(g)}\) is diagonal, the components of each \(\gamma^{(g)}_{i}\) in \(Y^{(g)}\) are mutually independent and the thresholds corresponding to distinct rows are also conditionally independent. Hence, we need only derive the marginal conditional distributions for \(\alpha^{(g)}_{k}, \ k = 1, \ldots, s\). Let \(\{\gamma^{(g)}_{ki}\}\) be the order statistics of \(\{\gamma^{(g)}_{ki}\}\) such that \(\gamma^{(g)}_{k1} \leq \cdots \leq \gamma^{(g)}_{kn_{k}}\), and \(n^{(g)}_{ki}\) be the total number of \(\gamma^{(g)}_{ki}\) with \(\{\gamma^{(g)}_{ki} = t, i = 1, \ldots, n_{g}\}\) for \(t = 0, \ldots, b_{k}\). Among the \(n_{g}\) observations of \(\gamma^{(g)}_{ki}\), there are \(n^{(g)}_{ki}\) of them equal to \(t\) and falling in the interval \([\alpha^{(g)}_{kt}, \alpha^{(g)}_{kt+1}]\). It can be shown (see Appendix I) that the conditional distribution of \(\alpha^{(g)}_{kt}\) is the following uniform distribution:

\[
[\alpha^{(g)}_{kt}|X,Z,\theta,Y,F] \overset{D}{=} U[\gamma^{(g)}_{k,n^{(g)}_{k,0}} + \cdots + n^{(g)}_{k,t-1}, \gamma^{(g)}_{k,n^{(g)}_{k,0}} + \cdots + n^{(g)}_{k,t-1} + 1],
\]

for \(t = 2, \ldots, b_{k}\) and \(k = 1, \ldots, s\).

For structural parameters in \(\theta\), the estimation and the specification of the prior distributions are slightly different under distinct constraints (3), (4) and (5). Naturally, in estimating the unconstrained parameter, we need to specify its own prior distribution and the data in the corresponding group should be used. However, for constrained parameters across groups, only one prior distribution is necessary and all the data in the groups should be combined in the estimation. For notational simplicity, detailed discussions are first presented in the context of the constraints in (5). Based on the definition of the model and the nature of parameters, it is reasonable to assume the prior density of the components in \(\theta\) satisfies the following property (see also Arminger & Muthén, 1998):

\[
p(\Lambda, \Phi, \Psi) = p(\Lambda)p(\Phi)p(\Psi).
\]

Let \(\psi_{kk}\) be the \(k\)th diagonal element of \(\Psi\), and \(\Lambda_{k}\) be the \(k\)th row of \(\Lambda\); for any \(k \neq j\), we further assume the prior distributions of \(\psi_{kk}\) and \(\psi_{jj}\), as well as \(\Lambda_{k}\) and \(\Lambda_{j}\), are independent. According to the suggestions given in Broemeling (1985), Lee (1981), Lee & Zhu (2000) and Lindley & Smith (1972), the following conjugate prior distributions are used:

\[
p(\psi_{kk}^{-1}) \sim \text{Gamma}[\eta_{0k}, \beta_{0k}], \quad p(\Lambda_{k}) \sim N[\Omega_{0k}, H_{0k}], \quad p(\Phi^{-1}) \sim W_q[R_0, \rho_0],
\]

where \(W_q[\cdot, \cdot]\) is the Wishart distribution, and \(\eta_{0k}, \beta_{0k}, \rho_0, \Lambda_{0k}, H_{0k}, R_0\) are hyperparameters whose values are assumed to be given. It has been pointed out in the work cited above that the conjugate prior distributions are sufficiently flexible in most applications.

Let \(U^{(g)} = (X^{(g)}, Y^{(g)})^\prime\), and \(U^{(g)}_{k}\) be the \(k\)th row of \(U^{(g)}\). It can be shown (see Appendix I) that the conditional distributions of the parameters in \(\theta\) are respectively equal to:

\[
[U^{(g)}_{k}|X,Z,\psi_{kk},\Lambda_{k},Y,F] \overset{D}{=} N[\omega_{k}, \Omega_{k}],
\]

where \(\omega_{k} = \Omega_{k}[\sum_{g=1}^{G} \psi_{kk}^{-1} U^{(g)} U^{(g)} + H^{-1}_{0k} \Lambda_{0k}]\) and \(\Omega_{k} = (\sum_{g=1}^{G} \psi_{kk}^{-1} F^{(g)} F^{(g)} + H^{-1}_{0k})^{-1};\)

\[
[\psi_{kk}^{-1}|X,Z,\Lambda_{k},Y,F] \overset{D}{=} \text{Gamma}[\eta_{k}, \beta_{k}],
\]

where \(\eta_{k} = n/2 + \eta_{0k}\) and \(\beta_{k} = 2^{-1} \sum_{g=1}^{G} [U^{(g)}_{k} U^{(g)}_{k} - 2\Lambda^{'}(g) F^{(g)} F^{(g)} F^{(g)} \Lambda_{k}] + \beta_{0k};\) and

\[
[\Phi^{-1}|F] \overset{D}{=} W_q[R, n + \rho_0],
\]

where \(R = (\sum_{g=1}^{G} F^{(g)} F^{(g)} + R_{0}^{-1})^{-1}\).
The conditional distributions for $\Lambda$ are given for the cases where all its elements are unknown parameters. Conditional distributions with some fixed elements in $\Lambda$ can be obtained using the derivation given in Shi & Lee (2000). Moreover, conditional distributions of the structural parameters under other constraints can be obtained as above, with slight modifications. For example, under constraints in (4), the prior distributions for $\Lambda_k$ and $\Phi$ are the same as before, while $p(Y_{kk}^{(g)}) \sim \text{Gamma}(\eta_{0k}^{(g)}, \beta_{0k}^{(g)})$, for $g = 1, \ldots, G$. It can be shown that the conditional distributions of $\Lambda$ and $\Phi$ are the same as (10) and (12), while the conditional distribution of $Y_{kk}^{(g)}$ is given by

$$[Y_{kk}^{(g)} - 1, Y_{kk}^{(g)} - 2] \sim \text{Gamma}[\eta_{kk}^{(g)}, \beta_{kk}^{(g)}],$$

where $\eta_{kk}^{(g)} = \eta_{kk}^{(g)} + \frac{n}{2}$ and $\beta_{kk}^{(g)} = 2^{-1} + \frac{n}{2} - 2\Lambda_k^{(g)} F^{(g)} U_{kk}^{(g)} + \Lambda_k^{(g)} (F^{(g)} F^{(g)}) \Lambda_k^{(g)} + \beta_{0k}^{(g)}$.

We now discuss the conditional distribution relating to the factor scores. Since the data from different groups are independent, $F^{(g)}$ and $F^{(h)}$ are independent for $g \neq h$, and we only need to find $[F^{(g)} | X, Z, \theta, \alpha, Y]$, which is equivalent to $[F^{(g)} | X^{(g)}, Z^{(g)}, \theta^{(g)}, \alpha^{(g)}, Y^{(g)}]$. Moreover, since $\xi_{i}^{(g)}$ and $\xi_{j}^{(g)}$ are independent for $i \neq j$, we need only find the marginal conditional distributions $[\xi_{i}^{(g)} | X^{(g)}, Z^{(g)}, \theta^{(g)}, \alpha^{(g)}, Y^{(g)}]$, $i = 1, \ldots, n_g$. It can be shown (see Appendix I) that

$$[\xi_{i}^{(g)} | X, Z, \theta, \alpha, Y] \sim N(\pi_{i}^{(g)}, \Pi^{(g)}), \quad g = 1, \ldots, G,$$

where $\pi_{i}^{(g)} = \Pi^{(g)} \Lambda^{(g)} \Psi^{(g)} \psi_{k}^{(g)} = (x_{i}^{(g)}, y_{i}^{(g)})^{T}$, $\Pi^{(g)} = (\Phi^{(g)} - 1 + \Lambda^{(g)} \Psi^{(g)} - 1 \Lambda^{(g)})^{-1}$.

Since $\Psi^{(g)}$ is diagonal, $y_{i}^{(g)}$ and $y_{j}^{(g)}$ are mutually conditionally independent for $k = 1, \ldots, s$ and $i = 1, \ldots, n_g$ when $\theta^{(g)}$ and $F^{(g)}$ are given. Thus, we need only derive the marginal conditional distributions $[y_{ki}^{(g)} | X, Z, \theta, \alpha, F] = [y_{ki}^{(g)} | X^{(g)}, Z^{(g)}, \theta^{(g)}, \alpha^{(g)}, F^{(g)}]$, $g = 1, \ldots, G$, $k = 1, \ldots, s$, $i = 1, \ldots, n_g$. It can be shown (see Appendix I) that

$$[y_{ki}^{(g)} | X, Z, \theta, \alpha, F] \sim N(\Lambda_{2}^{(g)} \xi_{i}^{(g)}, \psi_{k+r,k+r}^{(g)} I_{B}(y_{ki}^{(g)}), \Pi^{(g)}),$$

where $\Lambda_{2}^{(g)}$ is the $k$th row of $\Lambda^{(g)}$, $\psi_{k+r,k+r}^{(g)}$ is the $(k + r)$th diagonal element of $\Psi^{(g)}$, and $I_{B}(y)$ is an indicator function which takes value 1 if $y \in B$ and 0 otherwise.

It can be seen from expressions (8) and (10)–(15) that the conditional distributions required in implementing the Gibbs sampler are the familiar uniform, gamma, normal, univariate truncated normal and Wishart distributions. As expected, the algorithm based on these standard conditional distributions is efficient. Let $\{\theta_r, \alpha_r, Y_r, F_r, r = 1, \ldots, T\}$ be the random observations generated by the algorithm from the joint posterior distribution of $\alpha$, $\theta$, $F$ and $Y$ given ($X$, $Z$). Joint Bayesian estimates of $\alpha^{(g)}$, $\theta^{(g)}$ and $F^{(g)}$ are obtained as follows:

$$\alpha_k^{(g)} = \hat{E}(\alpha_k^{(g)} | X, Z) = \frac{1}{T} \sum_{t=1}^{T} \alpha_k^{(g)}, \quad k = 1, \ldots, s,$$

$$\hat{F}^{(g)} = \hat{E}(F^{(g)} | X, Z) = \frac{1}{T} \sum_{t=1}^{T} F^{(g)}, \quad (16)$$

$$\hat{\theta}^{(g)} = \hat{E}(\theta^{(g)} | X, Z) = \frac{1}{T} \sum_{t=1}^{T} \theta^{(g)}.$$

It follows from Geyer (1992) that these joint Bayesian estimates tend to their corresponding
posterior means in probability as $T$ tends to infinity. Posterior covariance matrices for revealing the dispersions of the estimates can also be estimated consistently from the generated observations. For example, the posterior covariance matrix of $\hat{\theta}^{(g)}$ is estimated by

$$
\hat{V}(\hat{\theta}^{(g)}|\mathbf{X}, \mathbf{Z}) = \frac{1}{T-1} \sum_{t=1}^{T} (\hat{\theta}^{(g)}_t - \hat{\theta}^{(g)})(\hat{\theta}^{(g)}_t - \hat{\theta}^{(g)})^T.
$$

(17)

Based on the simulated observations from the joint posterior distribution, we can also perform other statistical inferences in addition to point estimation. Examples are residual and outlier analyses and model diagnoses. To keep the paper to a reasonable length, these topics are not included.

4. Hypothesis testing and model comparisons

Further statistical inferences after estimation include testing of various null hypotheses about the model. For multi-sample models, it is particularly important to derive statistics for testing hypotheses on various relations among the structural parameters across groups. Following the suggestions of Berger (1985), Raftery (1993) and Kass & Raftery (1995), we will apply the Bayes factor in deriving the test statistic.

4.1. Bayes factor

In general, suppose there are two hypotheses $H_1$ and $H_2$ proposed for a data set $D$; and under $H_a$, the data are related to the parameter vector $\theta_a$ by a distribution with probability density $p(D|\theta_a, H_a)$. Given prior probabilities $p(H_1)$ and $p(H_2) = 1 - p(H_1)$, the data produce the posterior probabilities $p(H_1|D)$ and $p(H_2|D) = 1 - p(H_1|D)$. From Bayes’ theorem, we obtain

$$
p(H_a|D) = \frac{p(D|H_a)p(H_a)}{p(D|H_1)p(H_1) + p(D|H_2)p(H_2)}, \quad a = 1, 2,
$$

so that

$$
\frac{p(H_1|D)}{p(H_2|D)} = \frac{p(D|H_1)}{p(D|H_2)} \frac{p(H_1)}{p(H_2)}.
$$

The Bayes factor is defined as

$$
B_{12} = \frac{p(D|H_1)}{p(D|H_2)}.
$$

(18)

Hence, the posterior odds are equal to the product of the Bayes factor and prior odds. The Bayes factor $B_{12}$ is a summary of the evidence provided by the data in favour of $H_1$, as opposed to $H_2$; it measures how well the model associated with $H_1$ predicts the data relative to the model associated with $H_2$. From Kass & Raftery (1995), its interpretation is as described in Table 1.

The key quantity in computing the Bayes factor is the marginal density

$$
p(D|H_a) = \int p(D|\theta_a, H_a)p(\theta_a|H_a)d\theta_a,
$$

(19)
where $p(\theta_a|H_a)$ is the prior density of $\theta_a$ and $p(D|\theta_a, H_a)$ is the probability density (likelihood) of $D$ given $\theta_a$, under $H_a$. For some simple cases, this integral may be evaluated analytically. More often it is intractable, and hence searching for good approximations to this marginal density represents a challenging problem in the field; see Newton & Raftery (1994), Kass & Raftery (1995), Chib (1995), and DiCiccio et al. (1997), among many others.

One rough method for approximating the Bayes factor is the Schwarz (1978) criterion,

$$ S = \log p(D|\hat{\theta}_1, H_1) - \log p(D|\hat{\theta}_2, H_2) - \frac{1}{2}(d_1 - d_2) \log n, \quad (20) $$

where $\hat{\theta}_a$ is the maximum likelihood estimate of $\theta_a$ under $H_a$, $d_a$ is the dimension of $\theta_a$, and $n$ is the sample size. As $n$ tends to infinity, it has been shown (see Schwarz, 1978) that

$$ \frac{S - \log B_{12}}{\log B_{12}} \to 0, $$

and hence $S$ may be viewed as an approximation to $\log B_{12}$. An equivalent statistic is the Bayesian information criterion (BIC), which equals to minus twice the Schwarz criterion. As stated in Kass & Raftery (1995), $S$ or BIC is an often-used reference procedure for scientific reporting. However, the approximation is of order $O(1)$ so, even for large samples, it does not give the correct value.

In the context of Bayesian estimation via Gibbs sampling, an approach for computing the marginal density has been developed by Chib (1995). This approach is more accurate than the Schwarz criterion (or BIC) in the Bayes factor approximation. For a given hypothesis $H$, its basic idea is to express the marginal density as

$$ p(D|H) = \frac{p(D|\theta, H)p(\theta|H)}{p(\theta|D, H)}, $$

which holds for any $\theta$. Hence, for a given $\theta^*$, for example the mean of the posterior distribution, if the posterior density estimate at $\theta^*$ is denoted by $\hat{p}(\theta^*|D, H)$, then the proposed estimate of the logarithm marginal density is

$$ \log \hat{p}(D|H) = \log \hat{p}(D|\theta^*, H) + \log p(\theta^*|H) - \log \hat{p}(\theta^*|D, H). \quad (21) $$

In the following, for simplicity of notation, we write $p(\theta^*|D)$ for $p(\theta^*|D, H)$ if the context is clear. As Chib (1995) noted, $p(\theta^*|D)$ can be estimated by a sequence of simulations followed by a sequence of lower-dimensional density estimates. For example, letting $\theta^* = (\theta_1^*, \ldots, \theta_K^*)$ be a partition of $\theta^*$ into $K$ blocks, it follows that

$$ p(\theta^*|D) = p(\theta_1^*|D)p(\theta_2^*|D, \theta_1^*) \cdots p(\theta_K^*|D, \theta_1^*, \ldots, \theta_{K-1}^*), \quad (22) $$

where the first term is the marginal ordinate, which can be estimated from draws of the initial
Gibbs runs in the estimation, and the typical term is the reduced conditional ordinate
\[ p(\theta_k^*|D, \theta_1^*, \ldots, \theta_{k-1}^*) \] . The latter is given by
\[ \int p(\theta_k^*|D, \theta_1^*, \ldots, \theta_{k-1}^*, \theta_l(l > k), A)dp(\theta_{k+1}, \ldots, \theta_K, A|D, \theta_1^*, \ldots, \theta_{k-1}^*), \quad (23) \]
where \( A \) contains the augmented random quantities, and \( p(\cdot|\cdot) \) is being used to denote density and distribution function interchangeably. To estimate this term, continue the sampling with the complete conditional densities of \( \{\theta_k, \ldots, \theta_K, A\} \), where in each of these full conditional densities, \( \theta_h(h \leq k - 1) \) is set equal to \( \theta_h^* \). In general, it is easier to estimate \( p(\theta_k^*|D, \theta_1^*, \ldots, \theta_{k-1}^*) \) than \( p(\theta^*|D) \). As pointed out by a reviewer, the estimation of these posterior ordinates will require a very large number of simulations from the Gibbs sampler to be reliable, and since it is very small, its logarithm will be very large and will have a major impact on \( \log \hat{p}(D|H) \). See Chib (1995) for more discussion on the features and implementation of this method.

4.2. Bayes factor for multi-sample factor analysis model with mixed type variables

In the context of single-group structural equation modelling with continuous data, Raftery (1993) gave a detailed discussion of the application of the BIC to model selection. This statistic has also been used routinely by Jedidi, Jagpal, DeSarbo & Wedel (1996) in analysing finite mixtures of structural equation models with continuous data. For the present Bayesian analysis of multi-sample factor analyses model with mixed-type variables, the application of the Bayes factor is more difficult. In general, the Bayes factor can be sensitive to the priors, and using non-informative priors on parameters involved in the hypothesis is problematic. In the estimation, we use non-informative priors for the threshold parameters because this is a simple and natural choice for the situation with little prior information on the thresholds. In hypothesis testing, if the prior distributions of the parameters involved in the hypothesis are improper, the Bayes factor may be problematic; see for example Spiegelhalter & Smith (1982) in the context of linear and log-linear models. However, since the interesting hypotheses relating to constraints (3), (4), and (5) only involve the structural parameters of the factor analysis model, the thresholds can be regarded as nuisance parameters. As pointed out by Kass & Raftery (1995), under mild regularity conditions, the choice of prior on the nuisance parameters does not greatly affect the results, since the discrepancy of the Bayes factors from different priors is \( O(n^{-1}) \). Therefore, the choice of the prior for \( \alpha \) in our analysis is acceptable. Two procedures, one based on the Schwarz criterion (or BIC) and the other based on the computation of the marginal densities via (21), will be developed to approximate the Bayes factor.

In testing a hypothesis \( H_1 \) against its alternative \( H_2 \), it can be seen from (20) that the Schwarz criterion (or BIC) depends on the observed data log-likelihood function at \( \hat{\theta}_a \), that involves intractable multiple integrals. Hence, we encounter two difficulties, one in obtaining the maximum likelihood estimate \( \hat{\theta}_a \) and the other in obtaining the likelihood \( p(D|\hat{\theta}_a, H_a) \). These computational difficulties are solved by the idea of data augmentation and the Gibbs sampler. The proposed procedure is implemented via the following steps:

1. Under \( H_2 \), after the convergence of the Gibbs sampler algorithm, draw an observation \( (\theta_t, \alpha_t, Y_t, F_t) \) from the joint posterior distribution \[ [\theta, \alpha, Y, F|X, Z], \] where
Thesecondordinate \( \{ \text{where each term on the right-hand side is in the form of (23) with the augmented latent random quantities} \} \)

For \( a = 1, 2 \), generate new observations \( \{ \theta_{t,a,m}, m = 1, \ldots, M \} \), using the Gibbs sampler, from the conditional distribution \( \{ \theta | X, Z, Y_t, F_t \} \) under \( H_a \). Note that this conditional distribution is not dependent on the thresholds.

Evaluate

\[
S^{(t)} = \max_{1 \leq m \leq M} \log p(X, Y_t | \theta_{t,1,m}, H_1) - \max_{1 \leq m \leq M} \log p(X, Y_t | \theta_{t,2,m}, H_2) - \frac{1}{2} (d_1 - d_2) \log n,
\]

and \( B_{12}^{(t)} = \exp \{ S^{(t)} \} \), where

\[
\log p(X, Y_t | \theta_{t,a,m}, H_k) = -\frac{1}{2} \left[ \sum_{g=1}^{G} n_g \log |\Sigma_{t,a,m}^{(g)}| + \text{tr} \left\{ \sum_{g=1}^{G} \Sigma_{t,a,m}^{(g)-1} \left( \sum_{i=1}^{n_g} u_{it}^{(g)} u_{it}^{(g)T} \right) \right\} + pn \log (2\pi) \right],
\]

with \( u_{it}^{(g)} = (x_{it}^{(g)}, y_{it}^{(g)}) \), \( \theta_{t,a,m} = (\theta_{t,a,m}^{(1)}, \ldots, \theta_{t,a,m}^{(G)}) \) and \( \Sigma_{t,a,m}^{(g)} = \Sigma(\theta_{t,a,m}^{(g)}) \).

4. Update \( t \), and repeat steps 1 to 3 a total of \( T^* \) times, to obtain \( B_{12}^{(t)}, t = 1, \ldots, T^* \).

The Bayes factor is approximated by \( \hat{B}_{12} = T^{*^{-1}} \sum_{t=1}^{T^*} B_{12}^{(t)} \).

It can be seen that the above procedure is not directly affected by the prior distribution of \( \theta \). However, it may be indirectly affected because \( Y_t \), \( F_t \) and \( \theta_{t,a,m} \) are simulated observations from the conditional distributions whose forms are related to the prior distributions of \( \theta \). Some empirical results will be presented in the next section to reveal the sensitivity of the proposed procedure to the prior distributions.

In the second procedure, which we call the MD procedure, an estimate of the posterior density \( \hat{p}(\theta^* | D, H) \) in (21) will be obtained with \( \theta^* = (\Lambda^*, \Phi^*, \Psi^*) \) taken to be the Bayesian estimate and observed data \( D = (X, Z) \). As a special case of (22),

\[
p(\Lambda^*, \Psi^*, \Phi^* | X, Z) = p(\Phi^* | X, Z) p(\Psi^* | X, Z, \Phi^*) p(\Lambda^* | X, Z, \Phi^*, \Psi^*),
\]

where each term on the right-hand side is in the form of (23) with the augmented latent random quantities \( A = (\alpha, Y, F) \). The first ordinate \( p(\Phi^* | X, Z) \) can be estimated by taking the ergodic average of the full conditional density with the posterior draws \( \{ \Lambda_j, \Psi_j, \alpha_j, Y_j, F_j, j = 1, \ldots, J \} \) from the Gibbs runs in the estimation, leading to the estimate

\[
\hat{p}(\Phi^* | X, Z) = \frac{1}{J} \sum_{j=1}^{J} p(\Phi^* | X, Z, \Lambda_j, \Psi_j, \alpha_j, Y_j, F_j).
\]

The second ordinate \( p(\Psi^* | X, Z, \Phi^*) \) can be estimated similarly by continuing sampling for an additional \( J \) iterations from the conditional distribution \( p(\Lambda, \alpha, Y, F | X, Z, \Phi^*) \) via the Gibbs sampler. Let the newly simulated observations be \( \{ \Lambda_j, \alpha_j, Y_j, F_j, j = 1, \ldots, J \} \). Then

\[
\hat{p}(\Psi^* | X, Z, \Phi^*) = \frac{1}{J} \sum_{j=1}^{J} p(\Psi^* | X, Z, \Phi^*, \Lambda_j, \alpha_j, Y_j, F_j).
\]

The third ordinate \( p(\Lambda^* | X, Z, \Phi^*, \Psi^*) \) can be estimated similarly by newly simulated observations \( \{ \alpha_j, Y_j, F_j, j = 1, \ldots, J \} \) from the further reduced conditional distribution...
A simulation study was conducted to demonstrate the accuracy of the Bayesian estimates. It was assumed that model (1) has the following structure:

\[
\Lambda^{(g)} = \begin{pmatrix}
0.8^* & 0^* & \lambda_{31}^{(g)} & \lambda_{41}^{(g)} & \lambda_{51}^{(g)} & 0^* & 0^* & 0^* \\
0^* & 0.8^* & 0^* & 0^* & 0^* & \lambda_{62}^{(g)} & \lambda_{72}^{(g)} & \lambda_{82}^{(g)}
\end{pmatrix},
\Phi^{(g)} = \begin{pmatrix}
\phi_{11}^{(g)} & \phi_{12}^{(g)} \\
\phi_{12}^{(g)} & \phi_{22}^{(g)}
\end{pmatrix},
\Psi^{(g)} = \text{diag}(\psi_{11}^{(g)}, \ldots, \psi_{88}^{(g)}), g = 1, 2;
\]

where elements with an asterisk are fixed parameters. The first six variables were taken to be continuous, while the last two were polytomous. To create data sets for the simulation study,
Bayesian estimation and test for factor analysis model in several populations

Type 2.

Type 1.

we drew random samples \( \{ \xi_i^{(g)} \} \) and \( \{ \epsilon_i^{(g)} \} \) from multivariate normal distributions with mean zero and covariance matrices \( \Phi^{(g)} \) and \( \Psi^{(g)} \). Then \( \{ \chi_i^{(g)}, y_i^{(g)} \} \) was produced from \( \xi_i^{(g)} \) and \( \epsilon_i^{(g)} \) via (1) with the true values of the structural parameters taken as follows: \( \lambda_{kj}^{(g)} = 0.8 \), \( \psi_{kk}^{(g)} = 0.36 \), \( \phi_{11}^{(g)} = 1.0 \), \( \phi_{12}^{(g)} = 0.6 \), for \( g = 1, 2 \), \( k = 1, \ldots, 8 \), \( j = 1, 2 \). Each continuous observation \( y_i^{(g)} \) was then transformed to a polytomous observation \( z_i^{(g)} \) via (2) with the true thresholds \( \alpha_{11}^{(g)} = \alpha_{22}^{(g)} = (-1.0, -0.6, 0.6, 1.0) \). Similarly to the method given by Shi & Lee (1998) in identifying the polytomous variables, the following thresholds were fixed at the pre-assigned values: \( \alpha_{11}^{(g)} = -1.0 \), \( \alpha_{14}^{(g)} = 1.0 \), \( \alpha_{11}^{(g)} = -1.0 \), \( \alpha_{24}^{(g)} = 1.0 \); \( g = 1, 2 \). The remaining thresholds were treated as unknown parameters. Two sample sizes with \( n_1 = n_2 = 300 \) and \( n_1 = n_2 = 500 \) were considered.

To compare the Bayesian estimates with respect to different prior information, two types of prior for structural parameters were considered:

**Type 1.** Non-informative prior: \( p(\Lambda^{(g)}) \propto \text{constant} \), \( p(\Phi^{(g)}) \propto |\Phi^{(g)}|^{-(q+1)/2} \), and \( p(\Psi^{(g)}) \propto |\Psi^{(g)}|^{-1} \). For constrained parameters across groups, only one prior distribution is required; for example, under constraints (5), we need only specify \( p(\Lambda) \propto \text{constant} \), \( p(\Phi) \propto |\Phi|^{-(q+1)/2} \) and \( p(\Psi) \propto |\Psi|^{-1} \).

**Type 2.** Conjugate prior distributions as given in Section 3 with the following hyperparameter values: \( \eta_{0k}^{(g)} = 10 \), \( \rho_0^{(g)} = 10 \), and \( H_{0k}^{(g)} = I \) (identity matrix of appropriate order). Since \( E(\Lambda_k) = \Lambda_{0k} \), \( E(\psi_{kk}) = \beta_{0k}(\eta_{0k} - 1) \), and \( E(\Phi) = R_0/(\rho_0 - q - 1) \), the others were chosen such that \( \Lambda_{0k}^{(g)} = \Lambda_{k+}^{(g)} \), \( \beta_{0k}^{(g)} = (\eta_{0k} - 1)\psi_{kk+}^{(g)} \) and \( R_0^{(g)} = (\rho_0^{(g)} - q - 1)\Phi_{++}^{(g)} \), where \( \Lambda_{k+}^{(g)} \), \( \psi_{kk+}^{(g)} \) and \( \Phi_{++}^{(g)} \) are the corresponding true population quantities. For constrained parameters across groups, only one prior distribution was specified.

Prior distributions corresponding to type 1 and type 2 respectively represent those with vague and good prior information. Joint Bayes estimates of the thresholds, the factor scores and the structural parameters subjected to constraints in (3), (4) and (5) were obtained via the proposed Gibbs sampler algorithm as described in Section 3. We first conducted a few test Gibbs runs to determine the burn-in phase of the Gibbs sampler algorithm, and observed that the algorithm converged rapidly in less than 1000 iterations with EPSR values less than 1.2 in all cases. Hence, we settled for a burn-in phase of 1000 Gibbs cycles and collected a total of \( T = 5000 \) additional observations to produce the joint Bayesian estimates of the parameters according to (16). Based on 100 replications, the mean and the root mean squares (RMS) between the estimates and the true values were computed. Results corresponding to estimates of the parameters with no constraints and with constraints in (5) under the two types of priors are given in Tables 2, 3 and 4, respectively. It can be seen from these tables that the Bayesian estimates are accurate with moderate sample sizes and are not sensitive to the prior distributions. To save space, factor scores estimates are not presented.

In contrast to estimation, the Bayes factor in hypotheses testing may be generally more sensitive to the choice of priors. Hence, another simulation study was conducted to reveal the behaviours of the Bayes factor in the analysis of the current multi-sample model with different degrees of prior information accuracy. The same specifications on the model as given above were used, except \( \Psi_{+}^{(1)} = 0.36I_8 \) and \( \Psi_{+}^{(2)} = 0.6I_8 \). The following
four hypotheses are considered:

\[ H_1: \text{with no constraints} \]
\[ H_2: \Lambda^{(1)} = \Lambda^{(2)}; \quad H_3: \Lambda^{(1)} = \Lambda^{(2)}, \Phi^{(1)} = \Phi^{(2)}; \]
\[ H_4: \Lambda^{(1)} = \Lambda^{(2)}, \Phi^{(1)} = \Phi^{(2)}, \Psi^{(1)} = \Psi^{(2)}. \]  

To give some empirical results to illustrate the sensitivity of the Bayes factor with respect to the prior distributions, the following different hyperparameter values were considered:

**Type I.** Same values as given in type 2 above, except \( \rho_{0k}^{(g)} = 2^{-1}(\eta_{0k} - 1)(\psi_{1k}^{(1)} + \psi_{2k}^{(2)}) \) under the constraint \( \Psi^{(1)} = \Psi^{(2)}. \)

**Type II.** \( \eta_{0k}^{(g)} = 10, \rho_{0k}^{(g)} = 10, \) with other hyperparameter values equal to twice those given in type I.

**Type III.** \( \eta_{0k}^{(g)} = 10, \rho_{0k}^{(g)} = 10, \) with other hyperparameter values equal to half those given in type I.

---

**Table 2.** Bayesian estimates of the parameters without constraints (\( n = 300 \))

<table>
<thead>
<tr>
<th>True parameter value</th>
<th>Non-informative prior</th>
<th>Conjugate prior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1</td>
<td>Group 2</td>
</tr>
<tr>
<td>( \lambda_{31}^{(g)} = 0.80 )</td>
<td>0.816</td>
<td>0.057</td>
</tr>
<tr>
<td>( \lambda_{41}^{(g)} = 0.80 )</td>
<td>0.814</td>
<td>0.057</td>
</tr>
<tr>
<td>( \lambda_{51}^{(g)} = 0.80 )</td>
<td>0.811</td>
<td>0.061</td>
</tr>
<tr>
<td>( \lambda_{62}^{(g)} = 0.80 )</td>
<td>0.828</td>
<td>0.080</td>
</tr>
<tr>
<td>( \lambda_{72}^{(g)} = 0.80 )</td>
<td>0.823</td>
<td>0.083</td>
</tr>
<tr>
<td>( \psi_{11}^{(g)} = 0.36 )</td>
<td>0.352</td>
<td>0.036</td>
</tr>
<tr>
<td>( \psi_{12}^{(g)} = 0.36 )</td>
<td>0.355</td>
<td>0.045</td>
</tr>
<tr>
<td>( \psi_{33}^{(g)} = 0.36 )</td>
<td>0.352</td>
<td>0.035</td>
</tr>
<tr>
<td>( \psi_{66}^{(g)} = 0.36 )</td>
<td>0.354</td>
<td>0.043</td>
</tr>
<tr>
<td>( \psi_{77}^{(g)} = 0.36 )</td>
<td>0.357</td>
<td>0.054</td>
</tr>
<tr>
<td>( \phi_{11}^{(g)} = 1.00 )</td>
<td>0.990</td>
<td>0.123</td>
</tr>
<tr>
<td>( \phi_{12}^{(g)} = 0.60 )</td>
<td>0.590</td>
<td>0.083</td>
</tr>
<tr>
<td>( \phi_{22}^{(g)} = 1.00 )</td>
<td>0.997</td>
<td>0.139</td>
</tr>
<tr>
<td>( \alpha_{12}^{(g)} = -0.60 )</td>
<td>-0.597</td>
<td>0.052</td>
</tr>
<tr>
<td>( \alpha_{13}^{(g)} = 0.60 )</td>
<td>0.588</td>
<td>0.055</td>
</tr>
<tr>
<td>( \alpha_{22}^{(g)} = -0.60 )</td>
<td>-0.604</td>
<td>0.051</td>
</tr>
<tr>
<td>( \alpha_{23}^{(g)} = 0.60 )</td>
<td>0.601</td>
<td>0.053</td>
</tr>
</tbody>
</table>
On the basis of these types of prior and the same sample sizes as before, the Bayes factors were computed using the Schwarz criterion (or BIC) and the MD approach as developed in Section 4.2. In using the Schwarz criterion, both $M$ and $T$ are taken to be 400. In the MD approach, $\theta^*$ was taken to be the corresponding constrained Bayesian estimate under the hypothesis obtained in the estimation. The numbers of observations generated in the computation of the posterior densities and the likelihood ratios were $J = 5000$ and $T = 5000$. The computed values of $2\log \hat{B}_{21}, 2\log \hat{B}_{32}$ and $2\log \hat{B}_{43}$ are reported in Table 5. From the values of $2\log \hat{B}_{21}$ and $2\log \hat{B}_{32}$, all the test statistics obtained from the different choices of prior information suggest the same conclusions that $H_2$ is more favourable than $H_1$, and $H_3$ is more favorable than $H_2$. These conclusions are in agreement with the true model. According to the values of $2\log \hat{B}_{43}$, the test statistics obtained from the above different settings also suggest the correct conclusion that $H_4$ should be rejected. We computed the posterior predictive $p$-values (Gelman et al., 1996) and found that the false model with constraints given in $H_4$ was not rejected.

In addition to the findings reported in Table 5, we have the following observations relating to the sensitivity of the Bayes factor estimate with respect to the hyperparameters: (i) The
MD approach gives close estimates of Bayes factor with different values of hyperparameters even with moderate sample size \( n_1 = n_2 = 300 \). This phenomenon is also true for the procedure based on the Schwarz criterion (or BIC). These results indicate that both approaches are not very sensitive to the prior information in testing hypotheses about the

<table>
<thead>
<tr>
<th>True value</th>
<th>Uniform prior</th>
<th>Conjugate prior</th>
<th>Uniform prior</th>
<th>Conjugate prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_{31} = 0.80 )</td>
<td>0.809 0.042</td>
<td>0.804 0.040</td>
<td>0.805 0.033</td>
<td>0.802 0.028</td>
</tr>
<tr>
<td>( \lambda_{41} = 0.80 )</td>
<td>0.806 0.046</td>
<td>0.807 0.042</td>
<td>0.803 0.032</td>
<td>0.803 0.031</td>
</tr>
<tr>
<td>( \lambda_{51} = 0.80 )</td>
<td>0.802 0.043</td>
<td>0.810 0.039</td>
<td>0.805 0.032</td>
<td>0.805 0.031</td>
</tr>
<tr>
<td>( \lambda_{62} = 0.80 )</td>
<td>0.801 0.041</td>
<td>0.806 0.040</td>
<td>0.801 0.317</td>
<td>0.802 0.030</td>
</tr>
<tr>
<td>( \lambda_{72} = 0.80 )</td>
<td>0.812 0.052</td>
<td>0.808 0.049</td>
<td>0.807 0.035</td>
<td>0.806 0.032</td>
</tr>
<tr>
<td>( \lambda_{82} = 0.80 )</td>
<td>0.812 0.048</td>
<td>0.803 0.045</td>
<td>0.807 0.037</td>
<td>0.807 0.033</td>
</tr>
<tr>
<td>( \psi_{11} = 0.36 )</td>
<td>0.353 0.025</td>
<td>0.360 0.026</td>
<td>0.357 0.021</td>
<td>0.360 0.020</td>
</tr>
<tr>
<td>( \psi_{22} = 0.36 )</td>
<td>0.356 0.030</td>
<td>0.361 0.027</td>
<td>0.357 0.024</td>
<td>0.360 0.022</td>
</tr>
<tr>
<td>( \psi_{33} = 0.36 )</td>
<td>0.355 0.027</td>
<td>0.357 0.025</td>
<td>0.360 0.021</td>
<td>0.361 0.021</td>
</tr>
<tr>
<td>( \psi_{44} = 0.36 )</td>
<td>0.356 0.025</td>
<td>0.366 0.026</td>
<td>0.357 0.024</td>
<td>0.358 0.019</td>
</tr>
<tr>
<td>( \psi_{55} = 0.36 )</td>
<td>0.358 0.026</td>
<td>0.359 0.025</td>
<td>0.359 0.025</td>
<td>0.361 0.020</td>
</tr>
<tr>
<td>( \psi_{66} = 0.36 )</td>
<td>0.357 0.032</td>
<td>0.360 0.026</td>
<td>0.361 0.024</td>
<td>0.361 0.023</td>
</tr>
<tr>
<td>( \psi_{77} = 0.36 )</td>
<td>0.358 0.038</td>
<td>0.353 0.034</td>
<td>0.360 0.028</td>
<td>0.360 0.025</td>
</tr>
<tr>
<td>( \psi_{88} = 0.36 )</td>
<td>0.360 0.034</td>
<td>0.362 0.031</td>
<td>0.360 0.031</td>
<td>0.361 0.026</td>
</tr>
<tr>
<td>( \phi_{11} = 1.00 )</td>
<td>1.002 0.099</td>
<td>0.992 0.087</td>
<td>1.001 0.078</td>
<td>1.000 0.070</td>
</tr>
<tr>
<td>( \phi_{12} = 0.60 )</td>
<td>0.598 0.055</td>
<td>0.599 0.060</td>
<td>0.601 0.046</td>
<td>0.595 0.047</td>
</tr>
<tr>
<td>( \phi_{22} = 1.00 )</td>
<td>1.007 0.083</td>
<td>0.993 0.080</td>
<td>1.007 0.071</td>
<td>0.994 0.072</td>
</tr>
<tr>
<td>( \alpha_{12}^{(1)} = 0.60 )</td>
<td>-0.599 0.053</td>
<td>-0.600 0.051</td>
<td>-0.604 0.043</td>
<td>-0.599 0.040</td>
</tr>
<tr>
<td>( \alpha_{13}^{(1)} = 0.60 )</td>
<td>0.588 0.055</td>
<td>0.597 0.054</td>
<td>0.607 0.042</td>
<td>0.603 0.038</td>
</tr>
<tr>
<td>( \alpha_{12}^{(2)} = -0.60 )</td>
<td>-0.604 0.053</td>
<td>-0.599 0.049</td>
<td>-0.606 0.043</td>
<td>-0.601 0.042</td>
</tr>
<tr>
<td>( \alpha_{13}^{(2)} = 0.60 )</td>
<td>-0.601 0.049</td>
<td>0.602 0.048</td>
<td>0.600 0.037</td>
<td>0.599 0.034</td>
</tr>
<tr>
<td>( \alpha_{22}^{(1)} = -0.60 )</td>
<td>-0.605 0.056</td>
<td>-0.601 0.052</td>
<td>-0.610 0.063</td>
<td>-0.605 0.048</td>
</tr>
<tr>
<td>( \alpha_{23}^{(2)} = 0.60 )</td>
<td>0.601 0.054</td>
<td>0.598 0.050</td>
<td>0.611 0.046</td>
<td>0.600 0.039</td>
</tr>
<tr>
<td>( \alpha_{22}^{(2)} = -0.60 )</td>
<td>-0.608 0.057</td>
<td>-0.603 0.052</td>
<td>-0.603 0.044</td>
<td>-0.604 0.035</td>
</tr>
<tr>
<td>( \alpha_{23}^{(2)} = 0.60 )</td>
<td>0.598 0.061</td>
<td>0.599 0.053</td>
<td>0.600 0.042</td>
<td>0.598 0.041</td>
</tr>
</tbody>
</table>

Table 4. Bayesian estimates of the parameters with constraints (5)

<table>
<thead>
<tr>
<th>Sample size, ( n )</th>
<th>Bayes factor</th>
<th>MD approach</th>
<th>Schwarz criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prior I</td>
<td>Prior II</td>
<td>Prior III</td>
</tr>
<tr>
<td>300</td>
<td>2 log ( B_{21} )</td>
<td>27.12</td>
<td>29.01</td>
</tr>
<tr>
<td></td>
<td>2 log ( B_{32} )</td>
<td>9.70</td>
<td>10.73</td>
</tr>
<tr>
<td></td>
<td>2 log ( B_{43} )</td>
<td>-17.95</td>
<td>-14.55</td>
</tr>
<tr>
<td>500</td>
<td>2 log ( B_{21} )</td>
<td>35.66</td>
<td>38.80</td>
</tr>
<tr>
<td></td>
<td>2 log ( B_{32} )</td>
<td>19.55</td>
<td>18.89</td>
</tr>
<tr>
<td></td>
<td>2 log ( B_{43} )</td>
<td>-20.45</td>
<td>-22.00</td>
</tr>
</tbody>
</table>
currently proposed model. (ii) The MD approach performs better than the Schwarz criterion in the sense that it gives stronger evidence to support the correct hypotheses. (iii) The differences between the two approaches decrease as the sample sizes increase.

5.2. Two real examples

In the first example, a small portion of the data set collected by the World Values Survey 1981–1984 and 1990–1993 (World Values Study Group, 1994) is analysed. The whole data set was collected in 45 societies around the world on broad topics such as work, the meaning and purpose of life, family life and contemporary social issues. As an illustration of our Bayesian approach, the data obtained from Canada (group 1, sample size \(n_1 = 479\)) and United Kingdom (group 2, sample size \(n_2 = 195\)) with a few selected variables were used.

Nine variables related to the respondents’ employment, homelife and religious belief were taken as manifest variables in \(u^{(g)} = (u_1^{(g)}, \ldots, u_g^{(g)})\), where \(u_3^{(g)}, u_7^{(g)}\) and \(u_9^{(g)}\) are polytomous variables with three, five and five categories, respectively; the other variables are continuous. Details of these variables are given in Appendix II. Based on some preliminary exploratory analyses, factor analysis models with three latent variables and the following specifications were proposed for \(g = 1, 2\);

\[
\Psi^{(g)} = \begin{pmatrix}
1 & \lambda_{21}^{(g)} & \lambda_{31}^{(g)} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & \lambda_{52}^{(g)} & \lambda_{62}^{(g)} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{83}^{(g)} & \lambda_{93}^{(g)}
\end{pmatrix}, \quad \Phi^{(g)} = \begin{pmatrix}
\phi_{11}^{(g)} & \phi_{12}^{(g)} & \phi_{13}^{(g)} \\
\phi_{21}^{(g)} & \phi_{22}^{(g)} & \phi_{23}^{(g)} \\
\phi_{31}^{(g)} & \phi_{32}^{(g)} & \phi_{33}^{(g)}
\end{pmatrix}
\]

and \(\Psi^{(g)}\) is a diagonal matrix, where zeros and ones stand for fixed parameter values. The thresholds \(\alpha_{11}^{(g)} > \cdots > \alpha_{13}^{(g)} > \alpha_{21}^{(g)} > \cdots > \alpha_{24}^{(g)} > \alpha_{31}^{(g)} > \alpha_{34}^{(g)}\), for \(g = 1, 2\), were fixed to some pre-assigned values to identify parameters involved in the polytomous variables. These fixed values were selected via \(\alpha_{kj}^{(g)} = \Phi^{-1}(f_k^{(g)})\), where the \(f_k^{(g)}\) are the observed cumulative marginal proportions of the categories with \(z_{kj}^{(g)} < j\), and \(\Phi^*\) is the distribution function of \(N(0,1)\). The latent variables can be roughly interpreted as job satisfaction (\(\xi_1\)), homelife (\(\xi_2\)) and religious belief (\(\xi_3\)).

In the Bayesian estimation, the values of the hyperparameters in the conjugate prior distributions for the structural parameters were selected as follows. We first conducted a Bayesian estimation for each group with non-informative prior distributions to obtain initial estimates \(\bar{\Lambda}^{(g)}, \bar{\Psi}^{(g)}\) and \(\bar{\Phi}^{(g)}\); then, we took \(\Lambda_0^{(g)} = \bar{\Lambda}^{(g)}, \beta_0^{(g)} = (\eta_0^{(g)} - 1)\bar{\Psi}_k^{(g)}\) and \(R_0^{(g)} = (\phi_0^{(g)} - q - 1)\bar{\Phi}^{(g)}\) as unconstrained parameters. As constrained parameters across groups we used the averages of the hyperparameters values over two corresponding groups. Finally, we took \(\eta_0^{(g)} = 10, \rho_0^{(g)} = 20\) and \(H_0^{(g)} = I\). Joint Bayesian estimates of the parameters subject to constraints as specified in (27) were obtained. The convergence of the Gibbs sampler was monitored by the EPSR values. To give some idea about the convergence, Fig. 1 presents the plots of the EPSR values against the iteration numbers for the case with no constraints among the parameters. It can be seen that the algorithm converged after about 2500 iterations. The convergences of the Gibbs sampler algorithm for the other cases are similar. After convergence, 5000 observations were collected to obtain the results. Bayesian estimates of the thresholds and the structural parameters without constraints are reported in Table 6, together with the standard error estimates. To save space, estimates subjected to other constraints are not presented.
We now discuss the hypotheses in (27). On the basis of the cultural background in Canada and United Kingdom, we expect the loading matrices and the covariance matrices of the three latent factors for job satisfaction, homelife and religious belief to be the same for these two countries. For the less interesting parameters relating to the variances of the error measurements, discrepancies between the two countries may or may not exist. Hence, the prior probabilities of hypotheses $H_2$ and $H_3$ concerning the equality of factor loadings and covariances are expected to be high. The prior probability of $H_1$ is expected to be low, and we do not have much insight into the prior probability of $H_4$. The Bayes factor is used to provide a formal test for these hypotheses. For the sake of comparison and completeness, both the Schwarz criterion (or BIC) and the MD approach are considered. The quantities $2 \log \hat{B}_{21}$, $2 \log \hat{B}_{32}$ and $2 \log \hat{B}_{43}$, computed via the Schwarz criterion with $M = T^* = 400$, are equal to 17.25, 10.43 and $-8.10$, respectively. The corresponding test statistics computed via the MD approach with $J = 5000$ and $T = 10000$ are equal to 14.15, 5.86 and $-13.10$, respectively. According to Table 1, these two methods lead to the same conclusion that the data give strong evidence to support $H_2$ and $H_3$ but not $H_4$. Hence, the constraints $\Lambda^{(1)} = \Lambda^{(2)}$ and $\Phi^{(1)} = \Phi^{(2)}$ for the more interesting structural parameters of the latent factors in the two groups are not rejected.

In the second example, four polytomous variables, each with three categories, were selected from a large set of life happiness and satisfaction items in the UCLA Longitudinal Study of Adolescent Growth (see Newcomb & Bentler, 1988). These four items are measures of feelings about: (i) sex life; (ii) relationship with partners or spouse; (iii) accomplishment in life; and (iv) overall satisfaction with work. Data were obtained from two groups of young adults: a female group with sample size 518, and a male group with sample size 221. Based
### Table 6. Bayesian estimates of parameters with no constraints: ICPSR data set

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Canada</th>
<th>United Kingdom</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>0.799</td>
<td>0.091</td>
</tr>
<tr>
<td>$\lambda_{31}$</td>
<td>-0.353</td>
<td>0.057</td>
</tr>
<tr>
<td>$\lambda_{52}$</td>
<td>1.251</td>
<td>0.109</td>
</tr>
<tr>
<td>$\lambda_{62}$</td>
<td>1.152</td>
<td>0.180</td>
</tr>
<tr>
<td>$\lambda_{83}$</td>
<td>3.417</td>
<td>0.246</td>
</tr>
<tr>
<td>$\lambda_{93}$</td>
<td>-1.080</td>
<td>0.095</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>0.799</td>
<td>0.151</td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>3.311</td>
<td>0.245</td>
</tr>
<tr>
<td>$\psi_{33}$</td>
<td>0.836</td>
<td>0.126</td>
</tr>
<tr>
<td>$\psi_{44}$</td>
<td>1.337</td>
<td>0.111</td>
</tr>
<tr>
<td>$\psi_{55}$</td>
<td>3.185</td>
<td>0.239</td>
</tr>
<tr>
<td>$\psi_{66}$</td>
<td>0.618</td>
<td>0.086</td>
</tr>
<tr>
<td>$\psi_{77}$</td>
<td>0.712</td>
<td>0.078</td>
</tr>
<tr>
<td>$\psi_{88}$</td>
<td>0.927</td>
<td>0.207</td>
</tr>
<tr>
<td>$\psi_{99}$</td>
<td>0.368</td>
<td>0.042</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>1.888</td>
<td>0.210</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>0.784</td>
<td>0.104</td>
</tr>
<tr>
<td>$\phi_{13}$</td>
<td>0.212</td>
<td>0.060</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>1.200</td>
<td>0.144</td>
</tr>
<tr>
<td>$\phi_{23}$</td>
<td>0.164</td>
<td>0.046</td>
</tr>
<tr>
<td>$\phi_{33}$</td>
<td>0.567</td>
<td>0.081</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>2.138</td>
<td>0.181</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-0.148</td>
<td>0.033</td>
</tr>
<tr>
<td>$\alpha_{32}$</td>
<td>0.356</td>
<td>0.039</td>
</tr>
<tr>
<td>$\alpha_{33}$</td>
<td>0.457</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>0.731</td>
<td>0.037</td>
</tr>
</tbody>
</table>

### Table 7. Bayes estimates of parameters: LA Longitudinal Study of Adolescent Growth

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Female group</th>
<th>Male group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>0.866</td>
<td>0.105</td>
</tr>
<tr>
<td>$\lambda_{42}$</td>
<td>0.515</td>
<td>0.103</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>0.310</td>
<td>0.056</td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>0.281</td>
<td>0.061</td>
</tr>
<tr>
<td>$\psi_{33}$</td>
<td>0.358</td>
<td>0.092</td>
</tr>
<tr>
<td>$\psi_{44}$</td>
<td>0.752</td>
<td>0.133</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>1.077</td>
<td>0.150</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>0.409</td>
<td>0.081</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>0.916</td>
<td>0.170</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>0.612</td>
<td>0.107</td>
</tr>
<tr>
<td>$\alpha_{32}$</td>
<td>1.071</td>
<td>0.078</td>
</tr>
<tr>
<td>$\alpha_{42}$</td>
<td>1.229</td>
<td>0.135</td>
</tr>
</tbody>
</table>
on the results given in a previous analysis (Lee et al., 1990), the data set was analysed by a confirmatory factor analysis model with the following specifications:

\[
\Lambda' = \begin{pmatrix}
1 & \lambda_{21} & 0 & 0 \\
0 & 0 & 1 & \lambda_{42}
\end{pmatrix}
\quad \text{and} \quad
\Phi = \begin{pmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{pmatrix},
\]

and a diagonal matrix \(\Psi\), where zeros and ones stand for fixed parameter values. Moreover, on the basis of their suggestion in identifying parameters in the polytomous variables, \(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{31}\) and \(\alpha_{41}\) were fixed at pre-assigned values. These pre-assigned values and the values of the hyperparameters were obtained via methods similar to those used in the first example. It was found that the Gibbs sampler algorithm in the Bayesian estimation with no constraints converged after about 3000 iterations, and the convergence behaviours in other cases were similar. To save space, the convergence summary is not displayed. After convergence, a total of 5000 observations were collected to obtain the estimates, which are reported in Table 7, together with the standard error estimates. Bayesian estimates under other constraints were also obtained similarly. We expect the prior probabilities of \(H_2, H_3\) and \(H_4\) to be higher than the prior probability of \(H_1\). Bayes factor estimates relating to the hypotheses given in (27) were again computed via the Schwarz criterion with \(M = T^* = 400\) and the MD approach with \(J = 5000\) and \(T = 10000\). With the MD approach, we found that \(2\log \hat{B}_{21} = 11.04, 2\log \hat{B}_{32} = 20.07\) and \(2\log \hat{B}_{43} = 3.11\); using the Schwarz criterion, the corresponding values were 14.33, 22.25 and 2.45, respectively. On the basis of the criterion given in Table 1, these results provide strong evidence to support the constraints \(\Lambda^{(1)} = \Lambda^{(2)}\) and \(\Phi^{(1)} = \Phi^{(2)}\). The constraint \(\Psi^{(1)} = \Psi^{(2)}\) is also not rejected, but the conclusion is not so definite. Hence, the difference in the covariance structures of the four polytomous variables between the female group and the male group is not significant.

6. Discussion

In this paper, the factor analysis model with mixed continuous and polytomous variables is generalized to permit simultaneous analysis of data from several groups or populations. This generalization has potential use in many practical applications. Using the basic procedures described in this paper, the development can be further generalized to the multi-sample LISREL model. One contribution in this paper is the development of a feasible estimation procedure for obtaining the Bayesian estimates of the thresholds, latent factor scores and structural parameters subjected to simple equality constraints. Another contribution is the development of statistics based on the Bayes factor for testing hypothesis of the model. Computation of the Bayes factor in the current complex model is non-trivial. Two procedures are developed: one based on the commonly used Schwarz criterion (or BIC), the other using Chib’s (1995) idea of computing the posterior densities and the Gibbs sampler to obtain the likelihood ratios. Results from empirical studies indicate that these procedures can be usefully applied to real studies.

In our Bayesian estimation, a non-informative prior is used for the nuisance threshold parameters, while the commonly used conjugate prior distributions are used for the structural parameters. In hypothesis testing, our empirical results roughly indicate that neither the Schwarz criterion (or BIC) nor the MD approach in computing the Bayes factors is very sensitive to the choices of hyperparameters in the conjugate prior distributions of the structural parameters. Of course, more detailed simulation studies are required to draw
more definite conclusions. Although our empirical results indicate that the Schwarz criterion (or BIC) is acceptable for scientific reporting, it is a rough approximation and not as good as the MD approach. Hence, we recommend use of the MD procedure.\textsuperscript{1} Since the hypotheses of interest just involve the structural parameters, the unknown thresholds are nuisance parameters. In general, it has been pointed out (by, for example, Kass & Raftery, 1995) that the choice of priors for the nuisance parameters does not greatly affect the results. In the context of our model, the improper prior for the nuisance thresholds has only an indirect effect on the Bayes factor through the conditional distributions involved in various Gibbs simulations. Hence, we expect that our results will conform to the general points made by Kass & Raftery (1995). More care should be taken in testing hypotheses involving the thresholds with improper priors. One simple method suggested in Kass & Raftery (1995) on the basis of an idea in Lempers (1971) is to set aside part of the data to use as a training sample which is combined with the non-informative prior distribution to produce an informative prior distribution. The Bayes factor is then computed from the remainder of the data. More advanced methods have recently been suggested; see, for example, O’Hagan (1991, 1995), Berger & Perrichi (1996) and Kass & Wasserman (1996). Application of these procedures to our model requires further research efforts.

The advantages of the Bayes factor approach have been well documented in the literature; see Kass & Raftery (1995) and the references therein. The chief limitations of the approach are its complexity in obtaining good approximations and its sensitivity to the priors and assumptions of the model. Recent developments that take advantage of advances in statistical computing have greatly improved the usefulness of the approach. Here, we demonstrate that it is a feasible method for analysing a complicated and useful model in psychometrics.

Acknowledgements

This research was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, project CUHK 4088/99H. The authors are greatly indebted to the Editor and anonymous reviewers for valuable comments in improving the paper; and also to P. M. Bentler, M. D. Newcombe, the ICPSR and the relevant funding agency for allowing the use of their data.

References


\textsuperscript{1} A computer program written in C for this procedure is available from the authors upon request. Software for Bayesian inference using Gibbs sampling (BUGS; see Gilks, Thomas & Spiegelhalter, 1994) is available from www.mrc-bsu.cam.ac.uk/bugs.


Appenedix I

Conditional distribution of $\alpha$ given $X$, $Z$, $\theta$, $Y$ and $F$

Using the notation given in Section 3.1,

$$p(\alpha^{(g)}_k | X, Z, \theta, Y, F) = p(\alpha^{(g)}_k | Z^{(g)}_k, Y^{(g)}_k, F^{(g)}_k, \theta^{(g)}_k)$$

$$= p(\alpha^{(g)}_k | Z^{(g)}_k, Y^{(g)}_k, F^{(g)}_k, \theta^{(g)}_k, \alpha^{(g)}_k) \propto c \prod_{i=1}^{n_g} p(z^{(g)}_{ki}, y^{(g)}_{ki} | F^{(g)}_k, \theta^{(g)}_k, \alpha^{(g)}_k),$$

$$p(z^{(g)}_{ki}, y^{(g)}_{ki} | F^{(g)}_k, \theta^{(g)}_k, \alpha^{(g)}_k) \begin{cases} \neq 0 & \text{if } y^{(g)}_{ki} \in \{\alpha^{(g)}_{k,z^{(g)}_{ki}}, \alpha^{(g)}_{k,z^{(g)}_{ki}+1}\} \\ = 0 & \text{otherwise} \end{cases}$$
This conditional density is non-zero in the region

\[ \Delta_k^{(g)} = \left\{ \begin{array}{l}
\tilde{y}_{k,n_{k,0}+n_{k,1}}^{(g)} < \alpha_{k,2}^{(g)} < \tilde{y}_{k,n_{k,0}+n_{k,1}+1}^{(g)} \\
\vdots \\
\tilde{y}_{k,n_{k,0}+\ldots+n_{k,b_k-1}}^{(g)} < \alpha_{k,b_k}^{(g)} < \tilde{y}_{k,n_{k,0}+\ldots+n_{k,b_k-1}+1}^{(g)}
\end{array} \right\}, \]

so

\[ p(\alpha_k^{(g)}) p(c_k^{(g)}, \tilde{y}_k) | F^{(g)}, \theta_k^{(g)}, \alpha_k^{(g)}) \propto \begin{cases} c_g^* & \text{if } \alpha_k^{(g)} \in \Delta_k^{(g)} \\ 0 & \text{otherwise,} \end{cases} \]

where \( c_g^* \) only depends on \( Z_k^{(g)}, Y_k^{(g)}, \theta_k^{(g)} \) and \( F^{(g)} \). Thus, the conditional distribution of \( \alpha_k^{(g)} \) is the following uniform distribution:

\[ [\alpha_k^{(g)} | X, Z, \theta, Y, F] \overset{D}{=} U[\tilde{y}_{k,n_{k,0}+\ldots+n_{k,i-1}}^{(g)}, \tilde{y}_{k,n_{k,0}+\ldots+n_{k,i-1}+1}^{(g)}], \quad t = 2, \ldots, b_k, \quad k = 1, \ldots, s. \]

**Conditional distribution of \( F^{(g)} \) given \( X, Z, \theta, \alpha \) and \( Y \)**

\[ p(F^{(g)} | X, Z, \alpha, \theta, Y) = p(F^{(g)} | X^{(g)}, Z^{(g)}, \alpha^{(g)}, \theta^{(g)}, Y^{(g)}) \]

\[ = p(F^{(g)} | U^{(g)}, \theta^{(g)}) = \prod_{i=1}^{n_g} p(\xi_i^{(g)} | u_i^{(g)}, \theta^{(g)}) \propto \prod_{i=1}^{n_g} p(\xi_i^{(g)} | \theta^{(g)}) p(u_i^{(g)} | \xi_i^{(g)}, \theta^{(g)}) \]

\[ \propto \prod_{i=1}^{n_g} \exp \left\{ -\frac{1}{2} \Phi_i^{(g)} \Phi_i^{(g)-1} \xi_i^{(g)} \right\} \times \exp \left\{ -\frac{1}{2} (u_i^{(g)} - \Lambda^{(g)} \xi_i^{(g)}) \Psi^{(g)-1} (u_i^{(g)} - \Lambda^{(g)} \xi_i^{(g)}) \right\} \]

\[ \propto \prod_{i=1}^{n_g} \exp \left\{ -\frac{1}{2} (\xi_i^{(g)} - \pi_i^{(g)}) \Pi^{(g)-1} (\xi_i^{(g)} - \pi_i^{(g)}) \right\}. \]

Hence,

\[ [\xi_i^{(g)} | X, Z, \theta, \alpha, Y] = [\xi_i^{(g)} | u_i^{(g)}, \theta^{(g)}) \overset{D}{=} N(\pi_i^{(g)}, \Pi^{(g)}), \quad i = 1, \ldots, n_g, \]

where

\[ \pi_i^{(g)} = \Pi^{(g)} \Phi_i^{(g)} \Psi^{(g)-1} u_i^{(g)}, \quad i = 1, \ldots, n_g, \quad \Pi^{(g)} = (\Phi^{(g)-1} + \Lambda^{(g)} \Psi^{(g)-1} \Lambda^{(g)-1})^{-1}. \]
Conditional distribution of $Y^{(g)}$ given $X$, $Z$, $\Theta$, $\alpha$ and $F$

$$p(Y^{(g)}|X, Z, \Theta, \alpha, F) = p(Y^{(g)}|Z^{(g)}, \alpha^{(g)}, \Theta^{(g)}, F^{(g)})$$

$$= \prod_{i=1}^{n} p(y_{i}^{(g)}|z_{i}^{(g)}, \alpha^{(g)}, \Theta^{(g)}, \xi_{i}^{(g)}) = \prod_{k=1}^{p} \prod_{i=1}^{n} p(y_{ki}^{(g)}|z_{ki}^{(g)}, \alpha_{k}^{(g)}, \Theta_{k}^{(g)}, \xi_{ki}^{(g)}).$$

From model (1) and with the notation of Section 3.1,

$$[y_{ki}^{(g)}|Z^{(g)}, F^{(g)}, \alpha^{(g)}, \Theta^{(g)}] \sim D N(\Lambda_{2k}^{(g)}\xi_{ki}^{(g)}, \psi_{k+r,k+r}^{(g)}I(\alpha_{k,2k}^{(g)}, \alpha_{k,2k+1}^{(g)})y_{ki}^{(g)}).$$

Conditional distribution of $\Lambda$ given $X$, $Z$, $\Psi$, $Y$ and $F$

$$p(\Lambda|X, Z, \Psi, Y, F) = \prod_{k=1}^{p} p(\Lambda_{k}|X_{k}, Y_{k}, \psi_{kk}, F) \propto \prod_{k=1}^{p} p(\Lambda_{k})p(X_{k}, Y_{k}|\Lambda_{k}, \psi_{kk}, F)$$

$$\propto \prod_{k=1}^{p} \exp \left\{ -\frac{1}{2} (\Lambda_{k} - \Lambda_{0k})^{T} H_{0k}^{-1} (\Lambda_{k} - \Lambda_{0k}) \right\}$$

$$\times \exp \left\{ -\frac{1}{2} \sum_{g=1}^{G} \sum_{i=1}^{n} \psi_{kk}^{-1} (u_{ik}^{(g)} - \Lambda_{k}^{i}\xi_{i}^{(g)})^{2} \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \sum_{k=1}^{p} \left[ (\Lambda_{k}^{T}H_{0k}^{-1}\Lambda_{k} - 2\Lambda_{k}^{T}H_{0k}^{-1}\Lambda_{0k}) + \sum_{g=1}^{G} \sum_{i=1}^{n} \psi_{kk}^{-1} (u_{ik}^{(g)} - \Lambda_{k}^{i}\xi_{i}^{(g)})^{2} \right] \right\}.$$

The above sum can be rewritten as

$$\Lambda_{k}^{T}H_{0k}^{-1}\Lambda_{k} - 2\Lambda_{k}^{T}H_{0k}^{-1}\Lambda_{0k} + \sum_{g=1}^{G} \psi_{kk}^{-1} U_{k}^{(g)} U_{k}^{(g)} - 2\Lambda_{k}^{T} \sum_{g=1}^{G} \psi_{kk}^{-1} F^{(g)} U_{k}^{(g)}$$

$$+ \text{tr} \left[ \Lambda_{k}^{T} \sum_{g=1}^{G} \psi_{kk}^{-1} F^{(g)} F^{(g)^{T}} \right] = \Lambda_{k}^{T} \left( H_{0k}^{-1} + \sum_{g=1}^{G} \psi_{kk}^{-1} F^{(g)} F^{(g)^{T}} \right) \Lambda_{k}$$

$$- 2\Lambda_{k}^{T} \left( H_{0k}^{-1}\Lambda_{0k} + \sum_{g=1}^{G} \psi_{kk}^{-1} F^{(g)} U_{k}^{(g)^{T}} \right) + \sum_{g=1}^{G} \psi_{kk}^{-1} U_{k}^{(g)} U_{k}^{(g)^{T}}.$$

Thus,

$$[\Lambda_{k}|X_{k}, Y_{k}, \psi_{kk}, F] \sim D N(\Omega_{k}, \Omega_{k}),$$

where $\omega_{k} = \Omega_{k}(\sum_{g=1}^{G} \psi_{kk}^{-1} F^{(g)} U_{k}^{(g)^{T}} + H_{0k}^{-1}\Lambda_{0k})$ and $\Omega_{k} = (\sum_{g=1}^{G} \psi_{kk}^{-1} F^{(g)} F^{(g)^{T}} + H_{0k}^{-1})^{-1}$. 
Conditional distribution of $\Psi$ given $X, Z, \Lambda, Y$ and $F$

$$p(\Psi|X, Z, \Lambda, Y, F) = \prod_{k=1}^{p} p(\psi_{kk}|X_k, Y_k, \Lambda_k, F)$$

$$\propto \prod_{k=1}^{p} p(\psi_{kk})p(U_k|\Lambda_k, \psi_{kk}, F) \propto \prod_{k=1}^{p} \psi_{kk}^{-(\eta_0k-1)} \exp(-\beta_0k\psi_{kk}^{-1}) \times \psi_{kk}^{-n/2}$$

$$\times \exp\left\{-\frac{1}{2} \sum_{g=1}^{G} \sum_{i=1}^{n_y} (u_{ik}^{(g)} - \Lambda_k \xi_i^{(g)}) \psi_{kk}^{-1} (u_{ik}^{(g)} - \Lambda_k \xi_i^{(g)}) \right\} \propto \prod_{k=1}^{p} (\psi_{kk}^{-1})^{n/2 + \eta_0k - 1}$$

$$\times \exp\left\{-\psi_{kk}^{-1}\left(\frac{1}{2} \sum_{g=1}^{G} [U_k^{(g)}U_k^{(g)\prime} - 2\Lambda_k F^{(g)}U_k^{(g)} + \Lambda_k'(F^{(g)\prime}F^{(g)})\Lambda_k] + \beta_0k\right)\right\}.$$ 

Thus,

$$[\psi^{-1}_{kk}|X_k, Y_k, \Lambda_k, F] \overset{D}{=} \text{Gamma}(\eta_k, \beta_k),$$

where $\eta_k = n/2 + \eta_0k$ and $\beta_k = 2^{-1} \sum_{g=1}^{G} [U_k^{(g)}U_k^{(g)\prime} - 2\Lambda_k F^{(g)}U_k^{(g)} + \Lambda_k'(F^{(g)\prime}F^{(g)})\Lambda_k] + \beta_0k$.

Conditional distribution of $\Phi$ given $X, Z, Y$ and $F$

$$p(\Phi|X, Z, Y, F) = p(\Phi|F) \propto \sum_{g=1}^{G} p(\Phi)p(F^{(g)}|\Phi)$$

$$\propto \sum_{g=1}^{G} \left|\Phi\right|^{-(\rho_0+q+1)/2} \exp\left\{-\frac{1}{2} \text{tr}(R_0^{-1}\Phi^{-1})\right\}$$

$$\times \left|\Phi\right|^{-n_x/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n_y} \xi_i^{(g)\prime} \Phi^{-1} \xi_i^{(g)}\right\}$$

$$\propto \left|\Phi\right|^{-(n+\rho_0+q+1)/2} \exp\left\{-\frac{1}{2} \text{tr}\left[\sum_{g=1}^{G} F^{(g)}F^{(g)\prime} + R_0^{-1}\right]\Phi^{-1}\right\}.$$ 

Thus,

$$[\Phi^{-1}|F] \overset{D}{=} W_q(R, n+\rho_0),$$

where $R = (\sum_{g=1}^{G} F^{(g)}F^{(g)\prime} + R_0^{-1})^{-1}$.

Appendix II

The number of the variable corresponding to the original data set is given in parentheses at the end of each question.

$u_1$: Overall, how satisfied or dissatisfied are you with your job? (V116)

$u_2$: How free are you to make decisions in your job? (V117)
$u_3$: How much pride, if any, do you take in the work that you do? (V115)

$u_4$: Overall, how satisfied or dissatisfied are you with your home life? (V180)

$u_5$: How satisfied are you with the financial situation of your household? (V132)

$u_6$: All things considered, how satisfied are you with your life as a whole these days? (V96)

$u_7$: Thinking about your reasons for doing voluntary work, how important are religious beliefs in your own case? (V62)

$u_8$: How important is God in your life? (V176)

$u_9$: How often do you pray to God outside of religious services? (V179)