

# Markov chain Monte Carlo Estimation Method of Confirmatory Factor Analysis Model with Mixed Data

Thanoon Y. Thanoon<sup>1\*</sup>, Hasmek Antranik warttan<sup>2</sup> and Robiah Adnan<sup>3</sup>

<sup>1</sup>(Department of Business Management Techniques, Administrative Technical College, Northern Technical University, Mosul, Iraq), thanoon.younis@ntu.edu.iq

<sup>2</sup>(Department of Business Management Techniques, Administrative Technical College, Northern Technical University, Mosul, Iraq), Hasmek.krekorian@ntu.edu.iq

<sup>3</sup>(Department of Mathematical Sciences, Faculty of science, University Technology Malaysia, Johor, Malaysia), Robiaha@utm.my

\*Correspondence Author: Thanoon Y. Thanoon: thanoon.younis@ntu.edu.iq

**Abstract.** This paper provides a general overview of (Bayesian Confirmatory Factor Analysis) with mixed ordinal and binary data. Mixed variables with specific cut-points are used and the simulation (Gibbs sampling) of the Markov chain Monte Carlo (MCMC) as an estimation tool. The problem of qualitative data is handled using censoring methods with specific cut points. Some additional tools, which contain on the Bayesian estimator, standard deviations (SD), Markov chain error (MC error) and highest posterior density (HPD) interval, are interpreted. The developed approach is discussed with the findings derived from the OpenBUGS program using the information on the quality of life (QOL).

**Keywords:** Confirmatory factor analysis(CFA), Bayesian inference, Gibbs sampling, Mixed data.

## 1. Introduction

"Confirmatory factor analysis" (CFA) is a special case of factor analysis which is commonly used in social researches. Its investigation needs more unequivocal and points by point data than "exploratory factor analysis" (EFA). For instance, in the case of data from past investigations, logical speculations or situations, it needs to identify which variable loads on which factor. In this case, confine particular loadings to be equivalent to zero. As well as, accessible data identified with the relationship between factors can embed.

The investigator has more often than not data accessible from past examinations, so he can estimate factor loadings by making particular assumptions. Sometimes CFA takes after EFA by dispensing with loadings with low values.

The introduction of CFA is very extraordinary comparing with EFA because it is an approval strategy model. As an outcome, the attack of the model will be poorer since a few parameters will be settled or dispensed with. What is lost in the attack will be pick it up as an expansion in the degree of opportunities along these lines for the stinginess standard. It is a factual system for displaying a succession of related information to gauge the interrelationships among manifest and latent variables [1].

In the former years, many specializers recommended models that contain CFA. Some of these papers are proposed by Erosheva and Curtis, [2]; Najafabadi, et al., [3]; Merkle, [4]; Kupek, [5]; Karatza, [6]; Burgette and Reiter, [7]; Hoofs, et al., [8]; Yu, et al., [9]; Taylor, [10]. Its strategy is very important in many applications, for example, cross-cultural research. The quick development of CFA is back to the request of unobtrusive models and related factual strategies to tackle complex research issues in various fields.

The approach of the Bayesian idea has been produced with the Gibbs sampling algorithm [11], the censoring distribution and latent variables are dealt with as assumptive missing data. Conjugate priors are utilized for structural parameters and non-informative priors are for the cut points.

Bayesian approach is used in this research as a proposed method for analyzing CFA models with qualitative data. For model selection, the Deviance Information Criterion (DIC) (see [12]) is used. The major problem is to solve the issue of qualitative data (mixed data), CFA and Bayesian analysis is used to estimate the parameters and treat the censoring distribution as a missing data and merge them in the posterior analysis.

The article contains the following: Introduction of confirmatory factor analysis (CFA) in Section 2. Using Bayesian estimation for CFA in Section 3. The Bayesian model selection using DIC is discussed in Section 4. An example of real data is given in Section 5. Section 6 discusses the findings and it's interpretations. Section 7 will be for the conclusions.

## 2. The Model of Confirmatory Factor Analysis

"The confirmatory factor analysis" (CFA) model is a normal extension of the exploratory factor analysis (EFA) model. For example, the following CFA is considered [1].

$$\mathbf{x} = \mathbf{A}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \quad (1)$$

where  $\mathbf{x}_i$  ( $p \times 1$ ) a random vector of response variables,  $\mathbf{A}$  ( $p \times q$ ) is a factor loading matrix of,  $\boldsymbol{\xi}$  ( $q \times 1$ ) is a of latent factors vector,  $\boldsymbol{\varepsilon}$  ( $p \times 1$ ) is called a random vector of residuals. So,  $\boldsymbol{\xi}$  is then distributed as  $N[\mathbf{0}, \mathbf{I}]$ , Further,  $\boldsymbol{\varepsilon}$  is distributed as  $N[\mathbf{0}, \boldsymbol{\Psi}_\varepsilon]$ , where  $\boldsymbol{\Psi}_\varepsilon$  with diagonal elements  $\psi_{\varepsilon 1}, \dots, \psi_{\varepsilon p}$  is a diagonal matrix. It has also been determined that, in this case,  $\boldsymbol{\varepsilon}$  and  $\boldsymbol{\xi}$  are both independent as well.

The proposed CFA Can be used to treat massive completely different kinds of cases. Moreover with reference to CFA s, one should watch out to properly interpret the mean vector  $\boldsymbol{\mu}$ , particularly because it relates to  $\boldsymbol{\mu}$ .

In any case, to solve the issue of qualitative data, it should be assumed that  $\mathbf{y}_i$  ( $s \times 1$ ) is a sub-vector of the unobservable continuous data. This information is reflected within the observed mixed vector  $\mathbf{z}_i$ . At its most essential level, mixed variable  $\mathbf{z}_m$  can be introduced according to it's unobserved continuous random variable  $\mathbf{y}_m$  by: Such that it is also true that  $\{-\infty = \alpha_{m,1} < \alpha_{m,2} < \dots < \alpha_{m,b_m} < \alpha_{m,b_m+1} = \infty\}$  is the set of cut point's specification which define the given classes, and  $b_m$  for which  $b_m$  represents the number of selected cut points regarding the mixed variable  $\mathbf{z}_m$ .

It should be clear that the number of cut points is equivalent to every mixed variable.

### 3. Markov Chain Monte-Carlo Method for CFA

In the previously recognized model, let  $\theta$  be a vector of the unknown parameter. To be more precise, to let Bayesian estimates generated according to the Gibbs sampler, it allowed being a vector containing the complete set of distinct unknown parameters. For the mixed variables, let  $\alpha$  be the vector of cut points.

Moreover, the observed mixed data is denoted by  $Z = (z_1, \dots, z_N)$ , while for latent continuous measurement let be denoted by  $Y = (y_1, \dots, y_N)$  which are related to  $Z$  respectively. Thence in the posterior analysis, the increase of observed data is done by  $Y$ . After establishing  $y$ , all the data can be counted and treated as continuous the problem is going to be easier to tackle. Moreover, let on the variable  $\Omega = (\omega_1, \dots, \omega_N)$  be latent matrix. In addition to that realization that the observations of  $\Omega$  and measurement equations intensified into a regular simultaneous regression model.

Therefore, issues related to the model's more complex elements can be solved by increasing the data. Through posterior analysis ( $Z$ ), which is the set of data observed, can be completed by  $(Y, \Omega)$ . Moreover, reveal the joint posterior distribution  $[\theta, \alpha, Y, \Omega | Z]$ . It is possible to apply the Gibbs sampling algorithm to create a series of observations from the associated posterior joint distribution.

As an outcome, a set of standard inferences based on a simulated sample of observations may be applied via the Bayesian approach. In addition, using Gibbs sampling, the iteration technique can be used to create a sample observation from the following distributions: " $[\Omega | \theta, \alpha, Y, Z]$ ", " $[\alpha, Y | \theta, \Omega, Z]$ " and " $[\theta | \alpha, Y, \Omega, Z]$ ".

With the same strategy, it can decide the non-informative before finding out " $\alpha$ " so that the relevant prior distribution is comparable to the constant, which are used in previous cut point problems. The conditional distribution " $[\theta | \alpha, Y, \Omega, Z]$ " hence can be added into individual components with several different structural parameters, depending on the particular models. The components of the conditional distribution, known as " $[\theta | \alpha, Y, \Omega, Z]$ " applied to prior distributions are somewhat various and different. To start with, it is implicitly assumed that the previous distributions for the unconstrained parameters from different models would act independently. In addition, it becomes necessary to identify the specific value of its prior distribution while constructing an approximation for the unconstrained parameters and to outline the information belonging to the corresponding groupings in order to be completely implemented.

This section explains the Bayesian estimation and model choice in the case of using CFA with mixed data. The definition of data increase is explored with Markov Chain Monte Carlo (MCMC) tools to complete the general scheme[13].

The goal is to define how to analyze the preceding CFA models, in the context of the mixed variables, using the Bayesian approach. Understanding these principles will allow us to strategically augment the data as termed in the Bayesian estimation of the confirmatory factor analysis with mixed variables.

The data matrix  $[Z]$  is enhanced by the unobserved data  $[Y, \Omega]$  in the posterior analysis. Along these lines so as to characterize the Bayesian estimation of CFA, let  $Z = \{z_1, \dots, z_n\}$

to represent the observed data set of mixed variables, and  $\theta$  represents the unknown parameters vector.

The Bayesian estimate for  $\theta$ , the estimates of standard error can be derived respectively from the mean sample and variance matrices.

$$\hat{\theta} = N^{-1} \sum_{t=1}^N \theta^{(t)}, \quad \widehat{\text{var}}(\theta | Z) = (N-1)^{-1} \sum_{t=1}^N (\theta^{(t)} - \hat{\theta})(\theta^{(t)} - \hat{\theta})'. \quad (2)$$

The following well-known prior distributions of conjugate are used:

$$\begin{aligned} p(\lambda_k) &\sim N[\lambda_{0k}, \mathbf{H}_{0k}], \quad p(\lambda_{\xi k} | \psi_{\delta k}) \sim N[\lambda_{0\xi k}, \psi_{\delta k} \mathbf{H}_{0\xi k}], \\ p(\Phi^{-1}) &\sim W_q[\mathbf{R}_0, \rho_0], \quad p(\psi_{\delta k}^{-1}) \sim \text{Gamma}[\alpha_{0k}, \beta_{0k}] \end{aligned} \quad (3)$$

Respectively given the definition that  $\psi_{\delta k}$ , is the  $k^{\text{th}}$  diagonal element of  $\Psi_{\delta}$ ,  $\lambda_k'$  and  $\lambda_{\xi k}'$  are the  $k^{\text{th}}$  rows of  $\Lambda$  and  $\Lambda_{\xi}$ .  $\mathbf{H}_{0\mu} = \text{diag}(\sigma_{01}^2, \dots, \sigma_{0p}^2)$ , and  $\lambda_{0k}, \lambda_{0\xi k}, \alpha_{0k}, \beta_{0k}, \rho_0, \sigma_{0k}, \mathbf{H}_0, \mathbf{H}_{0\xi k}$ , and  $\mathbf{R}_0$  it supposed to be known, however, prior information is obtained via analysis of past data, theoretical consideration and causal observance.

As built up in crafted by Kass and Raftery [14], expected prior knowledge, as it is related to current models, is usually chosen for comfort only when specific data are not collected enough. This should be possible on the grounds that the impact these presumptions have on Bayesian estimations stays little, notwithstanding when a huge example measure is utilized. The outcomes are useful when attempting to utilize PC displaying with the Gibbs sampler [11], because it can simulate  $\alpha, \theta$  and  $\Omega$ , all from the conditional distribution.

Because it is capable of simulating  $\alpha, \theta$  and  $\Omega$ , from all conditional distribution.

Notwithstanding, due to the presence of mixed data, excessively unpredictable can be the associated conditional distributions, making it impossible to effectively extract or simulate information from them. The simulation process will start with the starting values and then simulate the first observation  $(\alpha^{(1)}, \theta^{(1)}, \Omega^{(1)}, Y^{(1)})$  until the  $m^{\text{th}}$  iteration of these observations to get on with the observation from the posterior distribution.

The series, after the  $m^{\text{th}}$  repetition, will give us  $(\alpha^{(m+1)}, \theta^{(m+1)}, \Omega^{(m+1)}, Y^{(m+1)})$ . However, the joint distribution  $(\alpha^{(m)}, \theta^{(m)}, \Omega^{(m)}, Y^{(m)})$  can be proven to move toward the joint posterior distribution  $[\alpha, \theta, \Omega, Y | Z]$  [15].

#### 4. Bayesian Model Selection

In this paper, DIC is used provided by OpenBUGS. It is equal to:

$$DIC = \overline{D} + p_D = D(\overline{\theta}) + 2p_D \quad (4)$$

Where

$\overline{D}$ , the posterior mean of the deviance,

$p_D$ , the effective number of parameters,

and  $D(\bar{\theta})$  is the point estimate of the deviance at the mean of the estimated parameters  $\theta$ .

The deviance is defined as

$$D(\theta) = -2 \log f(y / \theta) \quad (5)$$

$$\text{While } \bar{D} = -\frac{2}{N} \sum_{t=B+1}^N \log f(y / \theta^{(t)}) \quad (6)$$

The model with the smallest DIC is evaluated as better predict an imitated dataset of an indistinguishable structure form that as noted at the moment [6]. The DIC values corresponding to the CFA s with real data is produced by OpenBUGS program.

## 5. Real Data Example

The (QOL) data measures the quality of life and has an extraordinary incentive for clinical works and medical research, arranging and assessment of social insurance. It has been for the most acknowledged part, QOL is a multidimensional idea that is best assessed by various distinctive idle builds, for example, physical capacity, wellbeing status, mental status, and social connections [16].

The instrument in QOL are normally estimated on the mixed scale, first, (12) variables are ordinarily with three-to five-categories and second (12) variables are binary with two categories. It is considerable in investigations of discrete nature likewise draws (Fayer and Machin, [17]; Fayer and Hand, [18]). This instrument WHOQOL-100 [16] was built up to assess four latent factors.

**Table 1** Description of the data

"The initial seven things (Q1 to Q7) are planned to address physical health".
"The following six things (Q8 to Q13) are proposed to address mental health".
"The three things (Q14, Q15, Q16) that take after are for social connections".
"The last eight things (Q17 to Q24) are expected to address the condition, giving a sum of 24 things".

The majority of the things are estimated with a five-point scale

"(1 = 'not at all/very dissatisfied'; 2 = 'a little/dissatisfied'; 3 = 'moderate/neither'; 4 = 'very much/satisfied'; 5 = 'extremely/very satisfied')", and we recode the second (12) variables to binary (1 = satisfied; 2 = dissatisfied) to study the relationship between ordinal and binary data.

The sample size is extremely large for entire data. To outline the Bayesian techniques, investigated by a synthetic data set with sample size  $n = 338$ . Regarding these mixed data as originated from a continuous normal distribution isn't right. So this Bayesian approach considers the discrete idea of the information which is connected to dissect this mixed data.

Manifest 24 real data example related to four basic latent variables ( $\xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4}$ ) from CFAMs defined in Equation (3) are discussed. To illustrate the confirmatory factor analysis Bayesian method with mixed variables, a real data set is related to random vectors,

$\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{i24})'$ , let  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{i24})'$  be the random vector of latent continuous, which corresponds to the mixed variables  $z_{i1}, z_{i2}, \dots, z_{i24}$  where  $z_i, i = 1, \dots, n$  are mixed variables that are related to (4) latent variables  $\mathbf{w}_i = (\xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4})$ ,  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{i24})$ , with the values of the parameters in  $\boldsymbol{\Lambda} = (\lambda_1, \lambda_2, \dots, \lambda_{20})'$ . Using CFA model to assess the relationships of the latent variables  $\mathbf{w}_i = (\xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4})$

The prior inputs of the hyperparameter values are as follow:

1- "Prior I: Elements in  $\lambda_{0k}$  and  $\lambda_{0\xi k}$  in Equation (3) are series equal to the following optional values with initial values are equal to 1";

" $\mathbf{R}_0^{-1} = 8\boldsymbol{\Phi}$ ,  $\mathbf{H}_{0\alpha}, \mathbf{H}_{0\beta}$  and  $\mathbf{H}_{0\xi k}$  are used to be 0.25 times the identity matrices;  $\alpha_{0k} = 10$ ,  $\beta_{0k} = 8$ ,  $\rho_0 = 30$ ".

2- "Prior II: Elements in  $\lambda_{0k}$  and  $\lambda_{0\xi k}$  in Equation (3) are series equal to the following optional values with initial values are equal to 0.5";

" $\mathbf{R}_0^{-1} = 8\boldsymbol{\Phi}$ ,  $\mathbf{H}_{0\alpha}, \mathbf{H}_{0\beta}$  and  $\mathbf{H}_{0\xi k}$  are used to be 0.25 times the identity matrices;  $\alpha_{0k} = 10$ ,  $\beta_{0k} = 8$ ,  $\rho_0 = 30$ ".

To explain and analyze the data set, the quality of life data set with (n=338) was analyzed by OpenBUGS [19] to implement Bayesian estimates in the CFA model. For analyzing it is used the MCMC procedure, the data need more iterations to converge when comparing the Bayesian analyses of CFA with the data. The estimation of the parameters were obtained using Bayesian approach from T=10000 Iterations for censored distribution in CFA model.

## 6. Results and Discussion

In this section, the objective is to discuss the results that reveal the empirical performance of estimating parameter and model selection.

The Bayesian way was used to analyse CFA models for mixed data, which is the purpose of this research, applied using freely available statistical software OpenBUGS to analyze unknown parameters and model selection (DIC). So it is convenient to be applied to real data. But there are some restrictions on this approach. The most important is that the majority of data come from mixed is discrete due to the nature of social and behavioral sciences. While studying mixed data, the essential assumption in CFA that the data that came from a normal distribution is plainly damaged, so it is necessary to find an alternative solution to overcome this problem.

Hence, it's clear that regarding mixed factors as the continuous variable may lead to prompt incorrect conclusions (see Lee et al., [20]; Olsson, [21]). A good approach to evaluating these kinds of data is to consider it observations came from censoring distribution with specific cut points.

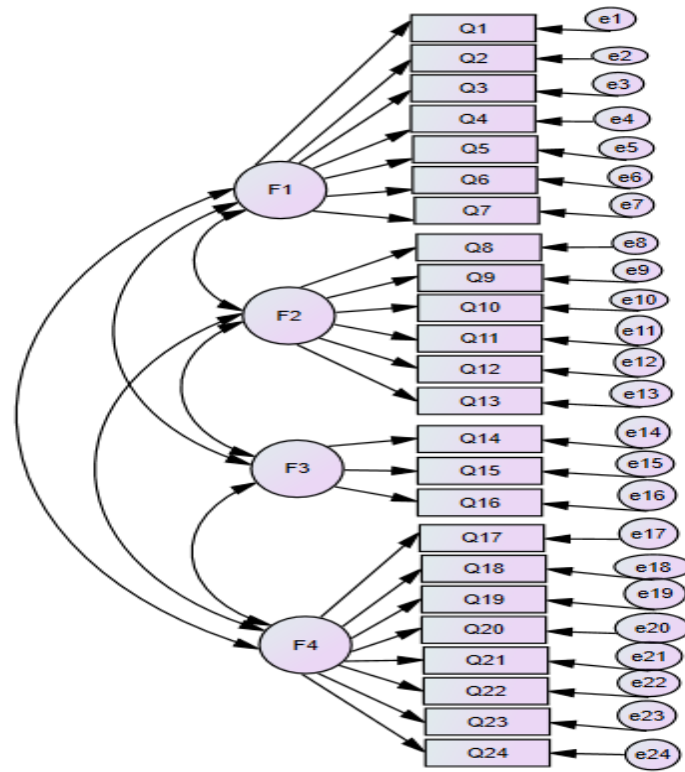
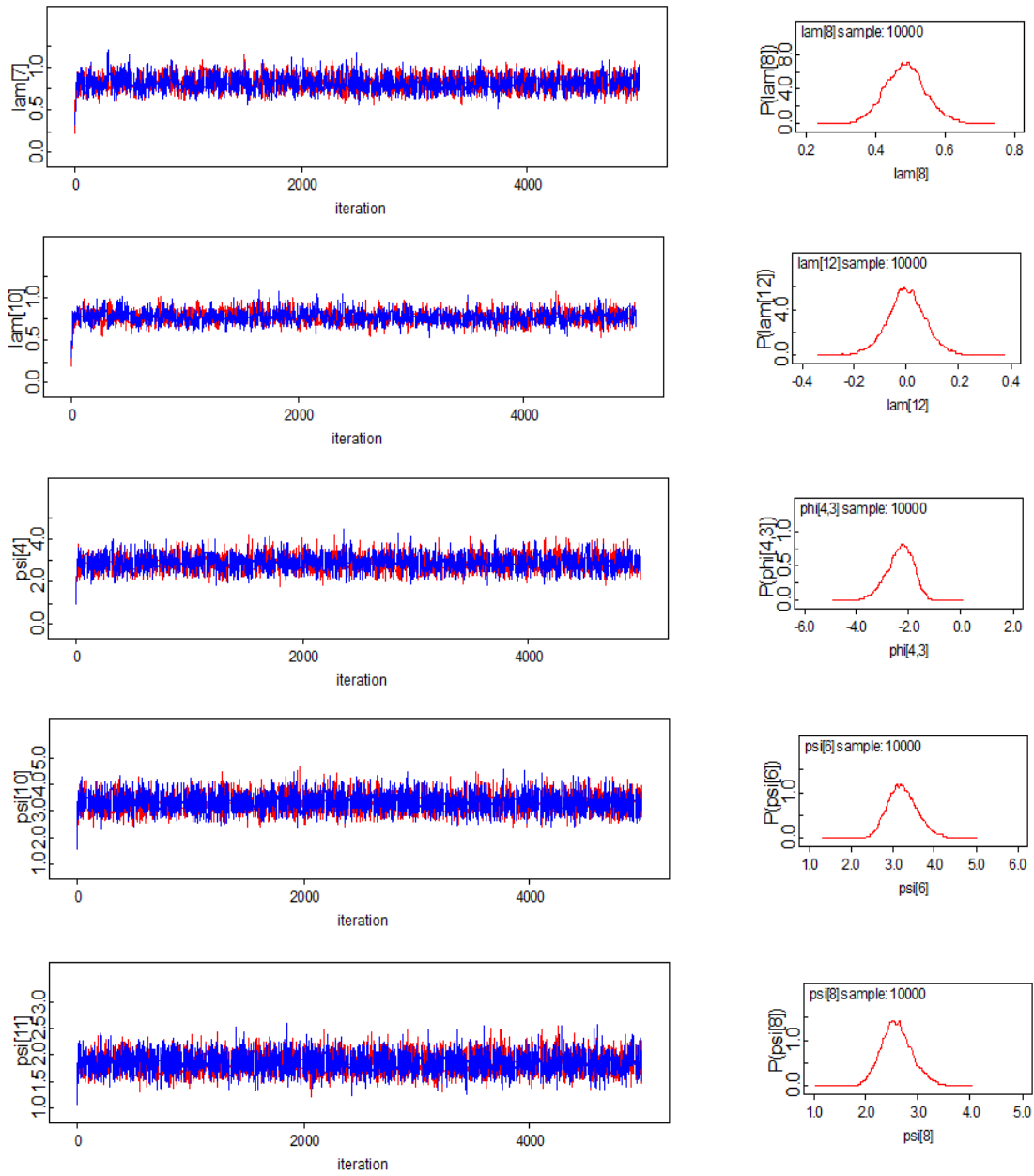


Figure 1. [Path Diagram of the study]



**Figure 2.** Time Series Plot and Posterior Density Plot of observation corresponding to some parameters for CFA with mixed Variables



**Table 2.** Bayesian Estimation of Confirmatory Factor Analysis Model with Mixed Variables

Para	Est.	SD	MC error	HPD Interval	Para	Est.	SD	MC error	HPD Interval
$\lambda_1$	0.906	0.122	0.004	[0.684, 1.157]	$\phi_{33}$	2.648	0.685	0.041	[1.481, 4.099]
$\lambda_2$	0.886	0.088	0.004	[0.725, 1.069]	$\phi_{34}$	-2.346	0.508	0.019	[-3.437, -1.461]
$\lambda_3$	0.829	0.080	0.003	[0.687, 0.993]	$\phi_{44}$	4.395	0.880	0.045	[2.860, 6.295]
$\lambda_4$	0.694	0.084	0.003	[0.535, 0.866]	$\Psi_1$	0.875	0.090	0.002	[0.710, 1.061]
$\lambda_5$	0.855	0.081	0.003	[0.711, 1.023]	$\Psi_2$	0.602	0.067	0.001	[0.480, 0.740]
$\lambda_6$	0.909	0.085	0.004	[0.755, 1.086]	$\Psi_3$	2.209	0.242	0.005	[1.771, 2.719]
$\lambda_7$	0.816	0.083	0.002	[0.660, 0.984]	$\Psi_4$	2.879	0.344	0.007	[2.259, 3.580]
$\lambda_8$	0.487	0.058	0.001	[0.375, 0.607]	$\Psi_5$	1.324	0.131	0.002	[1.080, 1.591]
$\lambda_9$	0.550	0.083	0.002	[0.393, 0.719]	$\Psi_6$	3.240	0.348	0.006	[2.624, 3.974]
$\lambda_{10}$	0.774	0.070	0.002	[0.643, 0.916]	$\Psi_7$	2.942	0.321	0.006	[2.362, 3.615]
$\lambda_{11}$	1.404	0.263	0.014	[0.962, 2.014]	$\Psi_8$	2.594	0.296	0.006	[2.064, 3.220]
$\lambda_{12}$	0.000	0.073	0.002	[-0.147, 0.145]	$\Psi_9$	2.088	0.218	0.004	[1.690, 2.536]
$\lambda_{13}$	1.210	0.283	0.019	[0.752, 1.830]	$\Psi_{10}$	3.320	0.312	0.004	[2.746, 3.968]
$\lambda_{14}$	0.467	0.135	0.006	[0.241, 0.764]	$\Psi_{11}$	1.850	0.176	0.003	[1.524, 2.211]
$\lambda_{15}$	0.406	0.127	0.006	[0.196, 0.689]	$\Psi_{12}$	3.628	0.376	0.006	[2.939, 4.397]
$\lambda_{16}$	0.857	0.199	0.011	[0.540, 1.320]	$\Psi_{13}$	1.389	0.401	0.021	[0.732, 2.270]
$\lambda_{17}$	0.336	0.119	0.005	[0.136, 0.608]	$\Psi_{14}$	1.744	0.422	0.020	[1.033, 2.688]
$\lambda_{18}$	1.399	0.290	0.017	[0.917, 2.045]	$\Psi_{15}$	1.119	0.343	0.016	[0.544, 1.881]
$\lambda_{19}$	1.228	0.259	0.016	[0.791, 1.782]	$\Psi_{16}$	1.392	0.413	0.022	[0.698, 2.334]
$\lambda_{20}$	1.384	0.257	0.016	[0.942, 1.960]	$\Psi_{17}$	0.737	0.216	0.012	[0.388, 1.217]
$\phi_{11}$	2.931	0.504	0.022	[2.071, 4.048]	$\Psi_{18}$	1.180	0.381	0.022	[0.566, 2.025]
$\phi_{12}$	-2.537	0.434	0.012	[-3.461, -1.780]	$\Psi_{19}$	1.154	0.358	0.018	[0.557, 1.940]
$\phi_{13}$	0.880	0.307	0.011	[0.330, 1.553]	$\Psi_{20}$	1.272	0.407	0.022	[0.610, 2.176]
$\phi_{14}$	-0.729	0.357	0.011	[-1.486, -0.067]	$\Psi_{21}$	1.167	0.350	0.017	[0.594, 1.926]
$\phi_{22}$	5.676	0.800	0.023	[4.303, 7.393]	$\Psi_{22}$	1.429	0.427	0.023	[0.718, 2.386]
$\phi_{23}$	-1.346	0.422	0.014	[-2.249, -0.580]	$\Psi_{23}$	1.389	0.412	0.023	[0.715, 2.311]
$\phi_{24}$	-0.404	0.488	0.016	[-1.385, 0.535]	$\Psi_{24}$	1.470	0.407	0.022	[0.785, 2.369]

**Table 3.** The goodness of Fit Statistics (DIC) for Confirmatory Factor Analysis Model with Mixed Variables

	<b>Dbar</b>	<b>Dhat</b>	<b>DIC</b>	<b>pD</b>
Total	13670.0	12810.0	14520.0	857.3

It is noticed from the table (2) that SD values are small for all the parameters and MC error, for mixed variables with censoring distribution and cut points. The effectiveness of the HPD intervals is suitable for mixed variables when using censored distribution, so the highest posterior density (HPD) for all the parameters has been calculated. Which are significantly far from zero.

The data sets are used confirmatory factor analysis to re-analyzed appear the effectiveness of DIC for the goodness of fit test and compare the results with the correct model, which is shown in table (3). The goodness of fit test DIC of CFA with mixed data using censored distribution was (14520.0). Which is considered a satisfactory performance of DIC.

By plotting several generated sequences of individual parameters with various starting values clear the idea of concourse Gibbs sampler, presented in figure Bayesian evaluations were acquired from T = 5000 cycles in CFA models for censoring distribution.

## 7. Conclusions and Recommendations

In social and behavioral sciences, the CFA is very common. The goals of this article are three, the first is to estimate parameters using confirmatory factor analysis models, the second is to deal with the problem of mixed data using censoring distribution and the thirds is applying the cut point's method. Developing a confirmatory factor analysis models with missing data may be a potential future research topic. There are examples with latent variables in CFA models, so this article discusses the measurable techniques acquire "SD" and Bayesian goodness of fit test using the (DIC) attributable and its complication. As we notice the troubles emerging caused by relationships between the latent factors and the discrete idea of mixed data are lightened by increasing data with some MCMC techniques. While increasing the actual observed data with the theoretical missing one would be moderately simple with the full information .This technique is intense and can be connected to other more unpredictable models.

## References

- [1] Lee, S.- Y. : Structural equation modeling: A Bayesian approach (Vol. 711). John Wiley & Sons. (2007).
- [2] Erosheva, E. A., & Curtis, S. M. : Dealing with rotational invariance in Bayesian confirmatory factor analysis. Technical Report 589, University of Washington. (2011).
- [3] Najafabadi, M. O., Hosseini, S. J. F., & Bahramnejad, S. : A Bayesian confirmatory factor analysis of precision agricultural challenges. African Journal of Agricultural Research, 6(5), 1219-1225. (2011).
- [4] Merkle, E. C. : A comparison of imputation methods for Bayesian factor analysis models. Journal of Educational and Behavioral Statistics, 36(2), 257-276. (2011).

- [5] Kupek, E. : Detection of Unknown Confounders by Bayesian Confirmatory Factor Analysis. *Advanced Studies in Medical Sciences*, 1(3), 143-56. (2013).
- [6] Karatza, A. S. : Bayesian Factor Analysis: Implementation on Schizotypal Personality Disorder Data (Doctoral dissertation, Athens University of Economics and Business). (2006).
- [7] Burgette, L. F., & Reiter, J. P. : Modeling adverse birth outcomes via confirmatory factor quantile regression. *Biometrics*, 68(1), 92-100. (2012).
- [8] Hoofs, H., van de Schoot, R., Jansen, N. W., & Kant, I. : Evaluating model fit in Bayesian confirmatory factor analysis with large samples: Simulation study introducing the BRMSEA. *Educational and psychological measurement*, 78(4), 537-568. (2018).
- [9] Yu, H., Campbell, M. T., Zhang, Q., Walia, H., & Morota, G. : Genomic Bayesian confirmatory factor analysis and Bayesian network to characterize a wide spectrum of rice phenotypes. *G3: Genes, Genomes, Genetics*, 9(6), 1975-1986. (2019).
- [10] Taylor, J. M. : Overview and Illustration of Bayesian Confirmatory Factor Analysis with Ordinal Indicators. *Practical Assessment, Research, and Evaluation*, 24(1), 4. (2019).
- [11] Geman, S. and Geman, D. : Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6), 721-741. (1984).
- [12] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. : Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64(4), 583-639. (2002).
- [13] Lee, S.-Y. and Song, X.-Y.: Basic and advanced structural equation models for medical and behavioural sciences. Hoboken: Wiley. (2012).
- [14] Kass, R. E. and Raftery, A. E.: Bayes factors. *Journal of the american statistical association*, 90(430), 773-795. (1995).
- [15] Geyer, C. J. : Practical markov chain monte carlo. *Statistical Science*, 473-483. (1992).
- [16] Power, M., Bullingen, M., Hazper, A. and WHOQOL Group: The World Health Organization WHOQOL-100: tests of the universality of quality of life in 15 different cultural groups worldwide. *Health Psychology*, 18, 495-505. (1999)
- [17] Fayer, P. M. and Machin, D.: Factor analysis. In M. J. Staquet, R. D. Hayes and P. M. Fayer (eds), *Quality of Life Assessment in Clinical Trials*. New York: Oxford University Press. (1998)
- [18] Fayer, P. M. and Hand, D. J. : Factor analysis, causal indicators and quality of life. *Quality of Life Research*, 8, 139-150. (1997)
- [19] Spiegelhalter, D., Thomas, A., Best, N. G. and Lunn, D.: *OpenBUGS user manual*. Medical Research Council Biostatistics Unit, Cambridge. (2007).
- [20] Lee, S.-Y., Poon, W.-Y. and Bentler, P.: Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics & probability letters*, 9(1), 91-97. (1990).
- [21] Olsson, U.: Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4), 443-460. (1979).