BAYESIAN APPROACH TO STRUCTURAL EQUATION MODELS FOR ORDERED CATEGORICAL AND DICHOTOMOUS DATA

THANOON Y. THANOON

A thesis submitted in fulfilment of the requirements for the award of the degree of Doctor of Philosophy (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

FEBRUARY 2017
To my beloved wife and my daughters Sara, Hajir and Maria
ACKNOWLEDGEMENT

First of all, my praises and thanks to Almighty Allah, the Most Gracious the Most Merciful, who gave me the knowledge, encouragement and patience to accomplish this research. I wish to express my sincere appreciation and thanks to my supervisor, Assoc. Prof. Dr. Robiah Adnan, for her encouragement when I have problems, guidance to the right way, and suggestions when I am hesitate, and critical reviews of this thesis. Without her continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to Universiti Teknologi Malaysia (UTM) to give me a chance to complete my PhD study and for the financial support for the Research University Grant (Q.J130000.2526.06H68) and the Ministry of Higher Education (MOHE) of Malaysia. I also need to take this opportunity to say thank you to the faculty of science staff for their guidance, advice and knowledge in this field to help me successfully finished my thesis. Without their contribution, interest and guidance I would not be able to complete this study.

I owe my loving thanks to my wife, who has given me unlimited support and love during my graduate study. I would also like to thank my beloved mother, father-in-law and mother-in-law, who have encouraged me to make my dreams come true. It is because of all of you that I am the person I have become. Finally, many thanks are due to everyone who helped me during my time at the Universiti Teknologi Malaysia. Many thanks to {Prof. Jichuan Wang} for advising the completion of this thesis and for providing the data.

I would also like to thank my other committee members for their time and effort. I am very thankful to all the staff in the department for their help. Last but not least, I want to thank my family for their encouragement and support throughout my graduate studies.
ABSTRACT

Structural equation modeling (SEM) is a statistical methodology that is commonly used to study the relationships between manifest variables and latent variables. In analysing ordered categorical and dichotomous data, the basic assumption in SEM that the variables come from a continuous normal distribution is clearly violated. A rigorous analysis that takes into account the discrete nature of the variables is therefore necessary. A better approach for assessing these kinds of discrete data is to treat them as observations that come from a hidden continuous normal distribution with a threshold specification. A censored normal distribution and truncated normal distribution, each includes interval, right and left where the later are with known parameters, are used to handle the problem of ordered categorical and dichotomous data in Bayesian non-linear SEMs. The truncated normal distribution is used to handle the problem of non-normal data (ordered categorical and dichotomous) in the covariates in the structural model. Two types of thresholds (having equal and unequal spaces) are used in this research. The Bayesian approach (Gibbs sampling method) is applied to estimate the parameters. SEM treats the latent variables as missing data, and imputes them as part of Markov chain Monte Carlo (MCMC) simulation results in the full posterior distribution using data augmentation. An example using simulation data, case study and bootstrapping method are presented to illustrate these methods. In addition to Bayesian estimation, this research provide the standard error estimates (SE), highest posterior density (HPD) intervals and a goodness-of-fit test using the Deviance Information Criterion (DIC) to compare with the proposed methods. Here, in terms of parameter estimation and goodness-of-fit statistics, it is found that the results with a censored normal distribution are better than the results with a truncated normal distribution, with equal and unequal spaces of thresholds. Furthermore, the results with unequal spaces of thresholds are less than the results of equal spaces of thresholds in the interval of the censored and truncated normal distributions, this is including the left censored and truncated normal distributions. The results of equal spaces of thresholds are less than the results of unequal spaces of thresholds in right censored and truncated normal distributions. In other cases, the results of bootstrapping method are better than the real data results in terms of SE and DIC. The results of convergence showed that dichotomous data needs more iterations to convergence than ordered categorical data.
Pemodelan persamaan struktur (SEM) adalah suatu kaedah statistik yang digunakan untuk mengkaji hubungan antara pembolehubah yang nyata dan pembolehubah terpendam. Dalam menganalisis data berkategorii turutan dan dikotomi, andaian asas dalam SEM bahawa pembolehubah datang dari taburan normal selanjau jelas bercanggah. Oleh itu satu analisis teliti yang mengambil kira sifat pembolehubah diskrit oleh itu adalah perlu. Pendekatan yang lebih baik untuk menilai jenis data diskret yang seperti ini supaya dapat menanganinya adalah dengan menganggapnya sebagai data yang datang dari taburan normal selanjau tersembunyi dengan nilai ambang yang spesifik. Taburan normal ditapis dan taburan normal terpangkas dengan selang kanan dan kiri, dimana yang kemudian adalah dengan parameter yang dikenali digunakan untuk menangan masalah kategori turutan dan dikotomi dalam SEM Bayes tidak linear. Taburan normal terpangkas tersebut digunakan untuk menangan masalah data yang tidak normal (data berkategorii turutan dan dikotomi) dalam covariates model struktur. Dua jenis nilai ambang (mempunyai ruang yang sama dan yang tidak sama) digunakan dalam kajian ini. Pendekatan Bayes (kaedah persampelan Gibbs) digunakan untuk menganggarkan parameter. SEM menangan pembolehubah terpendam sebagai data hilang, dan ia digunakan bagi melengkapkan keputusan simulasi Monte Carlo rantaian Markov (MCMC) untuk taburan posterior penuh dengan menggunakan pembesaran data. Satu contoh yang menggunakan data simulasi, kajian kes dan kaedah bootstrapping akan dibentangkan untuk menggambarkan keputusan dalam kajian ini. Sebagai tambahan kepada anggaran Bayes, kajian ini menyediakan anggaran ralat piawai (SE), selang ketumpatan posterior tertinggi (HPD) dan ujian kebagusan penyesuaian yang menggunakan Devians Kiteria Maklumat (DIC), digunakan sebagai perbandingan dengan keputusan dicadangkan. Dari segi anggaran parameter dan statistik kebagusan penyesuaian, didapati keputusan taburan normal ditapis adalah lebih baik daripada keputusan taburan normal terpangkas, bagi ruang yang sama dan tidak sama untuk nilai ambang. Tambahan pula, keputusan dengan ruang yang tidak sama bagi nilai ambang adalah kurang daripada keputusan untuk ruang sama bagi nilai ambang dalam selang untuk taburan normal ditapis dan dipangkas, ini termasuklah taburan normal ditapis kiri dan terpangkas. Keputusan bagi ruang sama untuk nilai ambang adalah kurang berbanding keputusan bagi ruang yang tidak sama untuk nilai ambang dalam taburan normal ditapis kanan dan terpangkas. Dalam kes lain, keputusan bagi kaedah bootstrapping adalah lebih baik daripada data sebenar dari segi SE dan DIC. Keputusan penumpuan menunjukkan bahawa data dikotomi memerlukan lebih banyak lelaran untuk menumpu berbanding data berkategorii turutan.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td>ii</td>
<td></td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iii</td>
<td></td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
<td></td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
<td></td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>vi</td>
<td></td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
<td></td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
<td></td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
<td></td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xvii</td>
<td></td>
</tr>
<tr>
<td>LIST OF SYMBOLS</td>
<td>xix</td>
<td></td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>xxi</td>
<td></td>
</tr>
</tbody>
</table>

1 INTRODUCTION 1
1.1 Problem Statement 3
1.2 Objectives of the Study 4
1.3 Scope of the Study 5
1.4 The Significance of the Research 5
1.5 The Organization of Thesis 6

2 LITERATURE REVIEW 8
2.1 The Introduction 8
2.2 Non-linear Structural Equation Models 12
2.3 Structural Equation Models for Ordered Categorical and Dichotomous Variables 13
2.4 The Bayesian Analysis of Structural Equation Models 17
2.5 Using Bayesian Analysis to Investigate Structural Equation Models with Dichotomous Variables 21
2.6 Using Bayesian Analysis to Investigate Structural Equation Models with Ordered Categorical Variables 22

2.7 Bayesian Analysis of Non-linear Structural Equation Models 23

2.8 Bootstrap Structural Equation Models 25

3 METHODOLOGY 29

3.1 General Definitions 30

3.1.1 Latent Variables 30

3.1.2 Covariates 30

3.1.3 Ordered Categorical Variables 31

3.1.4 Dichotomous Variables 32

3.2 Highest Posterior Density (HPD) 32

3.3 Gibbs Sampling Method and Markov Chain Monte Carlo (MCMC) 33

3.4 Understanding the Bayesian SEM with Ordered Categorical Variables and Covariates 34

3.4.1 Description of the Model 34

3.5 Using a Bayesian Analysis of Non-linear SEMs with Dichotomous Variables and Covariates 41

3.5.1 Description of The Model 41

3.6 Bayesian Estimation of SEMs Using Ordered Categorical and Dichotomous Variables and Covariates 43

3.7 Posterior Distribution Simulation to Approximate a Posterior Distribution 49

3.8 Features of the Prior Distribution 50

3.8.1 The Main Purposes of Priors 50

3.8.2 Setting Priors 51

3.9 Convergence of the Gibbs Sampling Algorithm 52

3.9.1 BGR DIAGNOSTIC 52

3.10 Proposed Methods 54

3.10.1 Introduction to Censoring 54

3.10.2 Introduction to Truncation 57

3.10.3 The Normal Distribution 59

3.10.4 The Truncated Normal Distribution 60

3.10.5 The Censored Normal Distribution 62

3.10.6 Similarities and Differences between Censoring and Truncation 63
3.10.7 Censoring and Truncation Using BUGS Programs 64
3.11 Data Augmentation 65
3.12 Model Comparison 66

4 SIMULATION STUDY 69
4.1 Overview 69
4.2 Sample Size for SEMs 70
4.3 Analysis of Simulated Data 71
4.4 The Results of Censored Normal Distribution 74
  4.4.1 Simulation Results and Discussion of BNSEMs for Ordered Categorical Data Using Censored Normal Distribution with Equal and Unequal Spaces of Thresholds as the Following 74
  4.4.2 Simulation Results and Discussion of BNSEMs for Dichotomous Data Using Censored Normal Distribution with Equal Spaces of Thresholds as the Following 86
4.5 The Results of Truncated Normal Distribution 93
  4.5.1 Simulation Results and Discussion of BNSEMs for Ordered Categorical Data Using Truncated Normal Distribution with Equal and Unequal Spaces of Thresholds as the Following 93
  4.5.2 Simulation Results and Discussion of BNSEMs for Dichotomous Data Using Truncated Normal Distribution with Equal Spaces of Thresholds 105
  4.5.3 Goodness of Fit Statistics Results for Censored Normal Distribution with Equal and Unequal Spaces of Thresholds 111
  4.5.4 Goodness of Fit Statistics Results for Truncated Normal Distribution with Equal and Unequal Spaces of Thresholds 113

5 A CASE STUDY 117
5.1 Real Example for Ordered Categorical Data 117
5.1.1 Data Description

5.2 Real Example for Dichotomous Data

5.2.1 Data Description

5.3 Bootstrapping method

5.3.1 Analysis of Real Data (Ordered Categorical)

5.3.2 Analysis of Real Data (Dichotomous)

5.4 Real Data Results and Discussion

5.4.1 Bayesian SEMs Results with Ordered Categorical Data Using Left Censored Normal Distribution with Unequal Spaces of Thresholds

5.4.2 Bayesian SEMs Results with Dichotomous Data Using Left Censored Normal Distribution with Equal Spaces of Thresholds

5.5 Goodness of Fit Statistics Results for left Censored Normal Distribution with Ordered Categorical Data and Unequal Spaces of Thresholds

5.6 Goodness of Fit Statistics Results for Left Censored Normal Distribution with Dichotomous Data and Equal Spaces of Thresholds

6 CONCLUSION AND FUTURE RESEARCH

6.1 Conclusion

6.2 Future Research

REFERENCES

APPENDIX A
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Cross-Tabulation of Two Ordered Categorical Variables X and Y</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>Cross-Tabulation of Two Dichotomous Variables X and Y</td>
<td>43</td>
</tr>
<tr>
<td>4.1</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Using Interval Censored Normal Distribution with Equal Spaces of Thresholds</td>
<td>74</td>
</tr>
<tr>
<td>4.2</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Using Interval Censored Normal Distribution with Unequal Spaces of Thresholds</td>
<td>75</td>
</tr>
<tr>
<td>4.3</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Using Right Censored Normal Distribution with Equal Spaces of Thresholds</td>
<td>78</td>
</tr>
<tr>
<td>4.4</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Using Left Censored Normal Distribution with Equal Spaces of Thresholds</td>
<td>79</td>
</tr>
<tr>
<td>4.5</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Using Right Censored Normal Distribution with Unequal Spaces of Thresholds</td>
<td>80</td>
</tr>
<tr>
<td>4.6</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Using Left Censored Normal Distribution with Unequal Spaces of Thresholds</td>
<td>81</td>
</tr>
<tr>
<td>4.7</td>
<td>Bayesian Estimation of NSEMs for Dichotomous Data Using</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interval Censored Normal Distribution with Equal Spaces of Thresholds</td>
<td>86</td>
</tr>
<tr>
<td>4.8</td>
<td>Bayesian Estimation of NSEMs for Dichotomous Data Using</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Right Censored Normal Distribution</td>
<td>89</td>
</tr>
<tr>
<td>4.9</td>
<td>Bayesian Estimation of NSEMs for Dichotomous Data Using</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Left Censored Normal Distribution</td>
<td>90</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>4.10</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data Using Interval Truncated Normal Distribution with Equal Spaces of Thresholds</td>
<td></td>
</tr>
<tr>
<td>4.11</td>
<td>Bayesian Estimation of Non-linear SEMs for Ordered Categorical Data Using Interval Truncated Normal Distribution with Unequal Spaces of Thresholds</td>
<td></td>
</tr>
<tr>
<td>4.12</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data Using Right Truncated Normal Distribution with Equal Spaces of Thresholds</td>
<td></td>
</tr>
<tr>
<td>4.13</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data Using Left Truncated Normal Distribution with Equal Spaces of Thresholds</td>
<td></td>
</tr>
<tr>
<td>4.14</td>
<td>Bayesian Estimation of NSEMs for Ordered Categorical Data Using Right Truncated Normal Distribution with Unequal Spaces of Thresholds</td>
<td></td>
</tr>
<tr>
<td>4.15</td>
<td>Bayesian Estimation of Nonlinear SEMs for Ordered Categorical Data Using Left Truncated Normal Distribution with Unequal Spaces of Thresholds</td>
<td></td>
</tr>
<tr>
<td>4.16</td>
<td>Bayesian Estimation of NSEMs for Dichotomous Data Using Interval Truncated Normal Distribution with Equal Spaces of Thresholds</td>
<td></td>
</tr>
<tr>
<td>4.17</td>
<td>Bayesian Estimation of NSEMs for Dichotomous Data Using Right Truncated Normal Distribution</td>
<td></td>
</tr>
<tr>
<td>4.18</td>
<td>Bayesian Estimation of NSEMs for Dichotomous Data Using Left Truncated Normal Distribution</td>
<td></td>
</tr>
<tr>
<td>4.19</td>
<td>Performance of Deviance Information Criterion (DIC) for BNSEMs with Equal Spaces of Thresholds Using Interval Censored Normal Distribution</td>
<td></td>
</tr>
<tr>
<td>4.20</td>
<td>Performance of Deviance Information Criterion (DIC) for BNSEMs with Equal Spaces of Thresholds Using Right Censored Normal Distribution</td>
<td></td>
</tr>
<tr>
<td>4.21</td>
<td>Performance of Deviance Information Criterion (DIC) for BNSEMs with Equal Spaces of Thresholds Using Left Censored Normal Distribution</td>
<td></td>
</tr>
<tr>
<td>4.22</td>
<td>Performance of Deviance Information Criterion (DIC) for BNSEMs with Unequal Spaces of Thresholds Using Interval Censored Normal Distribution</td>
<td></td>
</tr>
</tbody>
</table>
4.23 Performance of Deviance Information Criterion (DIC) for BNSEMs with Unequal Spaces of Thresholds Using Right Censored Normal Distribution

4.24 Performance of Deviance Information Criterion (DIC) for BNSEMs with Unequal Spaces of Thresholds Using Left Censored Normal Distribution

4.25 Performance of Deviance Information Criterion (DIC) for BNSEMs with Equal Spaces of Thresholds Using Interval Truncated Normal Distribution

4.26 Performance of Deviance Information Criterion (DIC) for BNSEMs with Equal Spaces of Thresholds Using Right Truncated Normal Distribution

4.27 Performance of Deviance Information Criterion (DIC) for BNSEMs with Equal Spaces of Thresholds Using Left Truncated Normal Distribution

4.28 Performance of Deviance Information Criterion (DIC) for BNSEMs with Unequal Spaces of Thresholds Using Interval Truncated Normal Distribution

4.29 Performance of Deviance Information Criterion (DIC) for BNSEMs with Unequal Spaces of Thresholds Using Right Truncated Normal Distribution

4.30 Performance of Deviance Information Criterion (DIC) for BNSEMs with Unequal Spaces of Thresholds Using Left Truncated Normal Distribution

5.1 Bayesian Estimation of NSEMs for Ordered Categorical Data Using Left Censored Normal Distribution with Unequal Spaces of Thresholds

5.2 Bayesian Estimation of NSEMs for Dichotomous Data Using Left Censored Normal Distribution with Equal Spaces of Thresholds

5.3 Performance of Deviance Information Criterion (DIC) for BNSEMs with Unequal Spaces of Thresholds Using Left Censored Normal Distribution

5.4 Performance of Deviance Information Criterion (DIC) for BNSEMs with Unequal Spaces of Thresholds Using Left Censored Normal Distribution
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Flow Chart of the Study</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Path Diagram of Non-linear Structural Model</td>
<td>37</td>
</tr>
<tr>
<td>3.2</td>
<td>The hidden continuous normal distribution with a thresholds</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>specification</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>The hidden continuous normal distribution with a single of</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>threshold fixed at zero</td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>The Normal Distribution with Truncation and Censoring</td>
<td>60</td>
</tr>
<tr>
<td>3.5</td>
<td>Latent, Censored, and Truncation Variables</td>
<td>64</td>
</tr>
<tr>
<td>3.6</td>
<td>Flow Chart of the Proposed Methods</td>
<td>68</td>
</tr>
<tr>
<td>4.1</td>
<td>The Path Diagram of Non-linear SEMs (M₄) for the Simulation Data</td>
<td>72</td>
</tr>
<tr>
<td>4.2</td>
<td>Plots of BGR Diagnostic of λ₁, λ₂, β₂, β₃, and γ₃ for M₄ with</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Ordered Categorical Data Using Interval Censoring with Equal Spaces</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of Thresholds and n=200</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>Plots of BGR Diagnostic of β₁, β₃, β₄, λ₂ and λ₄ for M₄ with</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Ordered Categorical Data Using Interval Censoring with Unequal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spaces of Thresholds and n=200</td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td>Plots of BGR Diagnostic of λ₃, λ₂, β₃, γ₁, and γ₃ for M₄ with</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>Ordered Categorical Data Using Right Censoring with Equal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spaces of Thresholds and n=200</td>
<td></td>
</tr>
<tr>
<td>4.5</td>
<td>Plots of BGR Diagnostic of λ₃, λ₄, γ₂, γ₃, and β₄ for M₄ with</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Ordered Categorical Data Using Left Censoring with Equal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spaces of Thresholds and n=200</td>
<td></td>
</tr>
<tr>
<td>4.6</td>
<td>Plots of BGR Diagnostic of β₂, γ₃, β₄, γ₂, and ψₖₜδ for M₄ with</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Ordered Categorical Data Using Right Censoring with Unequal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spaces of Thresholds and n=200</td>
<td></td>
</tr>
<tr>
<td>4.7</td>
<td>Plots of BGR Diagnostic of λ₁, γ₃, λ₄, β₂, and γ₂ for M₄ with</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Ordered Categorical Data Using Left Censoring with Unequal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spaces of Thresholds and n=200</td>
<td></td>
</tr>
</tbody>
</table>
4.8 Plots of BGR Diagnostic of $\lambda_3$, $\lambda_4$, $\lambda_2$, $\beta_2$, and $\gamma_1$ for $M_4$ with Dichotomous data Using Interval Censored Normal Distribution with Equal Spaces of Thresholds and $n=200$

4.9 Plots of BGR Diagnostic of $\lambda_1$, $\gamma_3$, $\lambda_3$, $\lambda_4$, and $\beta_2$ for $M_4$ with Dichotomous Data Using Right Censored Normal Distribution with Equal Spaces of Thresholds and $n=200$

4.10 Plots of BGR Diagnostic of $\beta_4$, $\gamma_1$, $\lambda_4$, $\lambda_6$, and $\gamma_3$ for $M_4$ with Dichotomous Data Left Censored Normal Distribution with Equal Spaces of Thresholds and $n=200$

4.11 Plots of BGR Diagnostic of $\gamma_1$, $\lambda_2$, $\lambda_1$, $\lambda_6$, and $\Phi_{22}$ for $M_4$ with Ordered Categorical Data Using Interval Truncation with Equal Spaces of Thresholds and $n=200$

4.12 Plots of BGR Diagnostic of $\lambda_5$, $\lambda_6$, $\lambda_1$, $\lambda_3$, and $\beta_4$ for $M_4$ with Ordered Categorical Data Using Interval Truncation with Unequal Spaces of Thresholds and $n=200$

4.13 Plots of BGR Diagnostic of $\lambda_4$, $\beta_4$, $\lambda_1$, $\lambda_5$, and $\gamma_3$ for $M_4$ with Ordered Categorical Data Using Right Truncation with Equal Spaces of Thresholds and $n=200$

4.14 Plots of BGR Diagnostic of $\lambda_3$, $\lambda_6$, $\lambda_5$, $\beta_4$, and $\gamma_1$ for $M_4$ with Ordered Categorical Data Using Left Truncation with Equal Spaces of Thresholds and $n=200$

4.15 Plots of BGR Diagnostic of $\lambda_2$, $\lambda_3$, $\lambda_4$, $\lambda_5$, and $\beta_4$ for $M_4$ with Ordered Categorical Data Using Right Truncation with Unequal Spaces of Thresholds and $n=200$

4.16 Plots of BGR Diagnostic of $\beta_4$, $\gamma_3$, $\lambda_3$, $\lambda_6$, and $\beta_2$ for $M_4$ with Ordered Categorical Data Using Left Truncation with Unequal Spaces of Thresholds and $n=200$

4.17 Plots of BGR Diagnostic of $\gamma_1$, $\lambda_2$, $\lambda_4$, $\beta_1$, and $\beta_2$ for $M_4$ with Dichotomous Data Using Interval Truncated Normal Distribution with Equal Spaces of Thresholds and $n=200$

4.18 Plots of BGR Diagnostic of $\lambda_2$, $\beta_2$, $\lambda_3$, $\lambda_4$, and $\beta_4$ for $M_4$ with Dichotomous Data Using Right Truncated Normal Distribution with Equal Spaces of Thresholds and $n=200$

4.19 Plots of BGR Diagnostic of $\lambda_1$, $\lambda_4$, $\lambda_2$, $\lambda_3$, and $\gamma_2$ for $M_4$ with Dichotomous Data Using Left Truncated Normal Distribution with Equal Spaces of Thresholds and $n=200$

5.1 The path diagram of non-linear SEMs ($M_4$) for Real Data (Dichotomous)
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>Plots of BGR Diagnostic of $\lambda_1$, $\lambda_2$, $\lambda_5$, $\lambda_9$, and $\lambda_{12}$ for $M_4$ with Real Data (Ordered Categorical) Using Left Censored Normal Distribution with Unequal Spaces of Thresholds</td>
<td>129</td>
</tr>
<tr>
<td>5.3</td>
<td>Plots of BGR Diagnostic of $\lambda_5$, $\lambda_7$, $\lambda_8$, $\lambda_{10}$, and $\beta_1$ for $M_4$ with bootstrapping (Ordered Categorical) Using Left Censored Normal Distribution with Unequal Spaces of Thresholds</td>
<td>130</td>
</tr>
<tr>
<td>5.4</td>
<td>Plots of BGR Diagnostic of $\lambda_3$, $\lambda_5$, $\beta_3$, $\beta_4$, and $\gamma_3$ for $M_4$ with Real data (Dichotomous) using Left Censored Normal Distribution and Equal Spaces of Thresholds</td>
<td>132</td>
</tr>
<tr>
<td>5.5</td>
<td>Plots of BGR Diagnostic of $\lambda_4$, $\lambda_5$, $\gamma_2$, $\gamma_3$, and $\beta_4$ for $M_4$ with bootstrapping Method for (Dichotomous Data) using Left Censored Normal Distribution and Equal Spaces of Thresholds</td>
<td>133</td>
</tr>
<tr>
<td>5.6</td>
<td>Path Diagram of Psychological Data with Non-linear SEM for Ordered Categorical Data ($M_4$)</td>
<td>136</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>SEM</td>
<td>Structural Equation Modeling</td>
<td></td>
</tr>
<tr>
<td>SEMs</td>
<td>Structural Equation Models</td>
<td></td>
</tr>
<tr>
<td>NSEMs</td>
<td>Nonlinear Structural Equation Models</td>
<td></td>
</tr>
<tr>
<td>BSEMs</td>
<td>Bayesian Structural Equation Models</td>
<td></td>
</tr>
<tr>
<td>BNSEMs</td>
<td>Bayesian Nonlinear Structural Equation Models</td>
<td></td>
</tr>
<tr>
<td>PP</td>
<td>Posterior Predictive</td>
<td></td>
</tr>
<tr>
<td>ESEM</td>
<td>Exploratory Structural Equation Modeling</td>
<td></td>
</tr>
<tr>
<td>FA</td>
<td>Factor Analysis</td>
<td></td>
</tr>
<tr>
<td>EFA</td>
<td>Exploratory Factor Analysis</td>
<td></td>
</tr>
<tr>
<td>CFA</td>
<td>Confirmatory Factor Analysis</td>
<td></td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov chain Monte Carlo</td>
<td></td>
</tr>
<tr>
<td>HPD</td>
<td>Highest Posterior Density</td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>Deviance Information Creterion</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike Information Creterion</td>
<td></td>
</tr>
<tr>
<td>BF</td>
<td>Bayes Factor</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian Information Creterion</td>
<td></td>
</tr>
<tr>
<td>WLS</td>
<td>Weighted Least Squares</td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
<td></td>
</tr>
<tr>
<td>WLSMV</td>
<td>Weighted Least Squares with Mean and Variance Adjusted</td>
<td></td>
</tr>
<tr>
<td>MLR</td>
<td>Robust Maximum Likelihood</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>Generalized Least Squares</td>
<td></td>
</tr>
<tr>
<td>ULS</td>
<td>Unweighted Least Squares</td>
<td></td>
</tr>
<tr>
<td>MLMV</td>
<td>Maximum Likelihood with Mean and Variance Adjusted</td>
<td></td>
</tr>
<tr>
<td>MLM</td>
<td>Maximum Likelihood with Mean Adjusted</td>
<td></td>
</tr>
<tr>
<td>ADF</td>
<td>Asymptotically Distribution Free</td>
<td></td>
</tr>
<tr>
<td>PLS</td>
<td>Partial Least Square</td>
<td></td>
</tr>
<tr>
<td>GLLAMM</td>
<td>Generalized Linear Latent and Mixed Models</td>
<td></td>
</tr>
<tr>
<td>GLMM</td>
<td>Generalized Linear Mixed Models</td>
<td></td>
</tr>
</tbody>
</table>
CVM – Continuous/ Categorical Variable Methodology
MCCFA – Multiple group Categorical Confirmatory Factor Analysis
IRT – Item Response Theory
DIF – Differential Item Functioning
Cat – LS – Categorical Least Squares
MP – Multivariate Probit
LISREL – Linear Structural Relations
EQS – Structural Equation Modeling Software
Mplus – Muthén and Muthén
AMOS – Analysis of Moment Structures
IID – Independent and Identically Distributed
BUGS – Bayesian inference using Gibbs sampling
DBRDA – Distance-based Redundancy Analysis
BGR – Brooks Gelman Rubin
PDF – Probability Distribution Function
CDF – Cumulative Distribution Function
JAGS – Just Another Gibbs Sampler
LIST OF SYMBOLS

- $y$ – Underlying continuous normal variable corresponds to ordered categorical data
- $u$ – Underlying continuous normal variable corresponds to dichotomous data
- $z$ – Manifest variable
- $c_i$ – Vector of linear fixed covariates
- $q$ – Underlying continuous normal covariate corresponds to ordered categorical data
- $j$ – Underlying continuous normal covariate corresponds to dichotomous data
- $x_i$ – Vector of covariates
- $\theta$ – (theta) Matrix of unknown parameters
- $\beta$ – (beta) Coefficient matrix relating covariates $x$ to $\eta$
- $w$ – (omega) Matrix of latent variables
- $\alpha$ – (alpha) Unknown thresholds
- $\Lambda$ – (lambda) Factor loadings matrix
- $A$ – Matrix of unknown parameters
- $B$ – Coefficient matrix relating $\eta$ to $\eta$ in the structural equation
- $\Gamma$ – (gamma) Coefficient matrix relating $\xi$ to $\eta$ in the structural equation
- $\Phi$ – (phi) Variance-covariance matrices of latent variables $\xi$
- $\Psi$ – (psi) Variance-covariance matrices of $\zeta$
- $\eta$ – (eta) Outcome (dependent) latent vector in the structural equation
- $\xi$ – (zeta) Explanatory (independent) latent vector in the structural equation
- $\varepsilon, \delta$ – (epsilon), (delta) Random vectors of measurement errors
- $\Psi_{\varepsilon}, \Psi_{\delta}$ – Diagonal covariance matrices of measurement errors, with diagonal elements $\psi_{\varepsilon k}$ and $\psi_{\delta k}$, respectively
- $\alpha_0, \beta_0$ – Hyperparameters in the Gamma distribution related to the prior distribution of $\Phi$
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_0, \rho_0$</td>
<td>Hyperparameters in the Wishart distribution related to the prior distribution of $\Phi$</td>
</tr>
<tr>
<td>$\Lambda_{0ak}, H_{0ak}$</td>
<td>Hyperparameters in the multivariate normal distribution related to the prior distribution of the kth row of $\Lambda_{0ak}$ in the measurement equation</td>
</tr>
<tr>
<td>$I$</td>
<td>Identity Matrix</td>
</tr>
<tr>
<td>$R$</td>
<td>Ratio $R$ (= pooled / within) is red - for plotting purposes the pooled and within interval widths</td>
</tr>
<tr>
<td>$WSS$</td>
<td>The mean of the variances within each sample (within-sample variability)</td>
</tr>
<tr>
<td>$BSS$</td>
<td>The variance of the posterior mean values over all generated samples/chains (between-sample variance)</td>
</tr>
<tr>
<td>$T$</td>
<td>The number of iterations kept in each sample/chain</td>
</tr>
<tr>
<td>$k$</td>
<td>The number of generated samples/chains</td>
</tr>
<tr>
<td>$V$</td>
<td>The pooled posterior variance</td>
</tr>
<tr>
<td>$d$</td>
<td>The estimated degrees of freedom for the pooled posterior variance estimate</td>
</tr>
<tr>
<td>$R_c$</td>
<td>Corrected version of $R$</td>
</tr>
<tr>
<td>$c$</td>
<td>Censoring point or truncation point</td>
</tr>
<tr>
<td>$G_j$</td>
<td>The censoring time</td>
</tr>
<tr>
<td>$T^*$</td>
<td>Random variable in left censored normal and Equal to $\max(Y, G_j)$</td>
</tr>
<tr>
<td>$H_i$</td>
<td>The chronological time</td>
</tr>
<tr>
<td>$R_i$</td>
<td>The left truncated</td>
</tr>
<tr>
<td>$M_k$</td>
<td>Competing model</td>
</tr>
</tbody>
</table>
**LIST OF APPENDICES**

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Published and Accepted Papers</td>
<td>152</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

This chapter contains a brief introduction to Bayesian structural equation models (BSEMs) with linear fixed covariates and exogenous latent variables in the measurement model and non-linear covariates and exogenous latent variables in the structural model. Ordered categorical and dichotomous data are used in this research. Analysing data necessitates a determination of the type of data being analysed. However, there are many types of distributions; the most basic distribution of the data is the normal distribution. The similarity between the distribution of the data and the distribution assumed in the analysis can affect the validity of the result. It is not a good idea to assume the distribution of the dependent variables and the covariates to be normal when using ordered categorical and dichotomous data. However, routinely treating these types of data as coming from a normal distribution may lead to erroneous results. Structural equation models (SEMs) always assume that the dependent variables are normally distributed. In addition, the ordered categorical and dichotomous variables and covariates are assumed to be normally distributed when, in fact, they are discrete variables and covariates. Finally, ordered categorical and dichotomous variables and covariates often display positive or negative skew, such that the frequency for low counts is higher than the frequency when the count level increases, or the frequency for low counts is lower than that for high counts. A more appropriate analysis includes the specification of a censored normal distribution or a truncated normal distribution with known parameters and thresholds specification, rather than a normal distribution.

There are two main types of structural equation models, the first, is classical structural equation models which include maximum likelihood method (ML) and weighted least square method (WLS). The second, is Bayesian structural equation models such as Gibbs sampling method. However, in this thesis, we will focus on Bayesian SEMs using Gibbs sampling method.
Ordered categorical data are those with more than two categories, while dichotomous data are those with only two categories. Two types of distributions that used in this research, hidden continuous normal distribution are proposed for handling the problem of non-normal data in variables and covariates (ordered categorical and dichotomous). The hidden continuous normal distribution considered in this research are censored normal distribution and truncated normal distribution with known parameters. The choice between the models depends on the parameter estimation results and the goodness-of-fit statistics, Deviance Information Criterion (DIC). In many applications, the ordered categorical and dichotomous response variable data are from a complete data set. In this research, the focus is on the unobserved data that are hidden continuous normal data (censored and truncated normal distribution data). There are many researches that are carried out using structural equation models with full data. Thus, the likelihood function for a censored or truncated structural equation model is not the same as the likelihood function of a structural equation model with full data.

As we have seen, there is a great deal of research about Bayesian structural equation models; we can also find studies about ordered categorical and dichotomous data. However, there has been no research on BSEMs with linear fixed covariates and latent variables in the measurement model and non-linear covariates and latent variables in the structural model, for ordered categorical and dichotomous variables and covariates. The response variables in a Bayesian structural equation model for ordered categorical and dichotomous variables and covariates with unobserved data, however, there is not sufficient information for the over-dispersion problem in a normal distribution in the BSEMs. We are interested in discussing models with this kind of data, since the dependent variables are ordered categorical or dichotomous variables and covariates. Thresholds (cut points) are used to change the ordered categorical and dichotomous data to the continuous normal data in the variables and covariates using underlying continuous normal distribution. Therefore, when talking about BSEMs with unobserved data, that means a hidden continuous normal distribution with censoring or truncation.

Real data usually does not completely satisfy the assumptions often made by researches which result in a dramatic effect on the quality of statistical analysis. The development of bootstrap estimation technique has become very competitive for improving the efficiency of parameters in structural equation models. Bootstrap is a method for assigning measures of accuracy to sample estimates. The techniques are more than 30 years old and were first introduced by Efron (1979). The basic
The idea of bootstrap method is to generate observations that are randomly drawn with replacement from the original data set. The set of the selected sub-samples are considered as bootstrap samples and can be used to estimate the parameters of structural equation models. The bootstrap method obtained from such procedure is called pair bootstrap. However, these bootstraps are computer intensive method that can replace the theoretical formulation. The attractive feature of bootstrap method is that it does not rely on the assumption of normal distribution and is capable of estimating the standard error of any complicated estimator without any theoretical calculations. The interesting properties of the bootstrap techniques have to be traded off with computation cost and time (Rana et al., 2012).

1.1 Problem Statement

Several models and methods have been proposed in the past for analysing BSEMs with ordered categorical and dichotomous data. For instance, Song and Lee (2006a) proposed Bayesian structural equation models with non-linear covariates and latent variables in the structural model, with mixed continuous and dichotomous data, and they treated the dichotomous data in the covariates as a continuous normal distribution. A truncated normal distribution with unknown parameters was used to handle the problem of dichotomous data. The posterior predictive (PP) p-value was used as a goodness-of-fit statistics for model comparison. Lee (2007) used a hidden continuous normal distribution (truncated normal distribution) with unknown parameters to handle the problem of ordered categorical data. Lunn et al. (2009) and Lunn et al. (2012) mentioned that the truncated normal distribution is to be used only if there are no unknown parameters or if a censored prior distribution has been specified for the parameters. However, the truncated normal distribution is used when the dependent variables contain observed data only or when the dependent variables contain some unobserved data with no unknown parameters. Lu et al. (2012) used continuous normal distribution as a proposed method to handle the problem of ordered categorical variables in BSEMs, with an application to behavioral finance. In this thesis, the first problem is how to handle the problem of non-normal data (ordered categorical variables) in Bayesian non-linear SEMs to the measurement model. The second problem is how to improve the performance of Bayesian non-linear SEMs when there are non-normal data (dichotomous) in the variables in the measurement model. The third problem is how to find Bayesian analysis of non-linear structural equation models when there are (ordered categorical and dichotomous data) in the covariates in the structural model.
1.2 Objectives of the Study

A number of studies have presented applications of the analysis of Bayesian structural equation models with non-normal data. Most of these researches have considered hidden continuous normal distribution (truncated normal distribution with unknown parameters) for the data. The aim of this study is to discuss the problems of ordered categorical and dichotomous data and also covariates in these models, which means that we are going to discuss Bayesian non-linear structural equation models (BNSEMs) as a tool for the analysis of ordered categorical and dichotomous data with unobserved data in variables and covariates. This study embarks on the following objectives:

1. To handle the problem of ordered categorical variables in the measurement model, using a hidden continuous normal distribution (interval censored normal distribution, right and left censoring) and (interval truncated normal distribution, right and left truncation) with two types of thresholds (with equal and unequal spaces).

2. To improve the performance of BNSEMs with dichotomous variables using a hidden continuous normal distribution (interval censored normal distribution, right and left censoring) and (interval truncated normal distribution, right and left truncation) with known parameters and one type of thresholds (with equal spaces).

3. To handle the problem of non-normal data (ordered categorical and dichotomous) for the covariates in the structural model using an interval truncated normal distribution, and right truncation and left truncation with two types of thresholds (with equal and unequal spaces) for ordered categorical data and one type of thresholds (with equal spaces) for dichotomous data.

4. To evaluate the performance of the proposed methods when dealing with ordered categorical and dichotomous variables and covariates, through simulation, a case study and bootstrapping method. Comparisons are made based on goodness-of-fit statistics using Deviance Information Criterion (DIC).
1.3 Scope of the Study

In this research, we discuss BNSEMs with non-linear covariates and latent variables for ordered categorical and dichotomous variables and covariates. The focus is on hidden continuous normal distribution methods, with both censored normal distribution (which includes interval, right and left censoring) and truncated normal distribution (which includes interval, right and left truncation) with known parameters, to handle the problem of ordered categorical and dichotomous variables. Furthermore, three different kinds of truncation – interval, right and left truncation with known parameters are going to handle the problem of discrete data in the covariates. The Gibbs sampling method is used to perform the parameter estimation for the proposed models. It will discuss the previous models in the simulation study and apply the proposed methods to simulated data by using the R-program, and analyse these using the R2OpenBUGS package in the R-program. First, the simulated data is made up of SEMs with ordered categorical variables and covariates. The second simulation involves data with SEMs and dichotomous variables and covariates. One sample size \( n=200 \), two different initial values and two different types of thresholds (with equal and unequal spaces) for ordered categorical data and one type of thresholds (with equal spaces) for dichotomous data will be considered, to investigate the effects of different sample sizes and different initial values on the proposed methods. The proposed methods (left censored normal distribution with unequally spaces of thresholds for ordered categorical data, and left censored normal distribution with equally spaces of thresholds for dichotomous data) will be applied to a case study and bootstrapping method (resampling the real data) with SEMs for ordered categorical and dichotomous variables and covariates with sample size \( n=200 \). The non-linear models are selected in this research using a combination between covariates and exogenous latent variables. Quadratic and cubic effects are used in the proposed models to explain the non-linear effects on the endogenous latent variables.

1.4 The Significance of the Research

Analysis of structural equation models is very important since these models are used in many fields, including scientific, social and behavioral sciences. Many researchers have proposed methods to obtain accurate estimates of parameters for different types of variables. This research focuses on the problem of ordered categorical and dichotomous data in variables and covariates in Bayesian non-linear structural equation models. The focus is to determine which of the proposed methods
is the best for estimating the parameters of the models. The best model can be identified by using the statistical error measurements such as standard error (SE) and highest posterior density (HPD). The Deviance Information Criterion (DIC) is used as a goodness-of-fit statistic to compare the performance of the proposed models. To handle the problems referred to above, we use a hidden continuous normal distribution (censored and truncated normal distribution) in Bayesian analysis with a Markov chain Monte Carlo (MCMC) simulation method for analysing non-linear structural equation models. In this approach, the standard error estimates (SE) can easily be obtained through simulated observations from the joint posterior distribution by the MCMC methods, and the Bayesian analysis can be conducted in SEM by using the OpenBUGS program and R2OpenBUGS package in the R-program. The Gibbs sampler algorithm is used to implement the Bayesian SEMs. The Bayesian SEM methodology permits the user to utilize the prior information for updating the current information on the parameter. This research provides the parameter estimation, standard error (SE), highest posterior density (HPD) and the goodness-of-fit statistics (DIC) of the proposed models, and these results can help researchers in determining the appropriate model with ordered categorical and dichotomous data in variables and covariates.

1.5 The Organization of Thesis

The goal of this research is to introduce some new methods to handle the problem of ordered categorical and dichotomous data in variables and covariates in Bayesian non-linear structural equation models. The proposed research work has been introduced, with an overview, a statement of the problem, objectives, scope and the significance of the research. The research study deals with the problem of ordered categorical and dichotomous variables and covariates in Bayesian structural equation models. To reach this goal, an improved approach was proposed that will enable us to handle the problem of these types of variables, based on the objectives that have been defined in this chapter.

The rest of the work is organized as follows: Chapter 2 reviews the relevant literature, including research related to structural equation models (SEMs), non-linear SEMs, Bayesian SEMs, classical and Bayesian SEMs with ordered categorical and dichotomous variables, Bayesian non-linear SEMs, and Bootstrap SEMs. In Chapter 3, we present our methodology, which uses Bayesian structural equation models with non-linear covariates and latent variables for ordered categorical and dichotomous data in variables and covariates. We also present some proposed methods. Chapter 4
through the T(,) construct, whose syntax is the same as that of the I(,) construct in OpenBUGS. However, the interpretation is different. The I(,) mechanism through which censored data is dealt with, cannot be used for modeling truncated distribution as long as they lack unknown parameters. In case there are parameters that are unknown, then there will be wrong inferences. Hence, the I(,) construct does not work well for truncated distribution where there are unknown parameters. Nonetheless, the I(,) construct is necessary for making truncated distribution with known parameters (Plummer, 2012).

3.11 Data Augmentation

Data augmentation, which occurs as the result of latent or auxiliary variables, is a method of converting the unobserved variable to more usable form, like usual linear regression. This allows the Gibbs sampler to be used, and simplifies calculation. One reason why data augmentation is used, is to ensure that the conjugacy of the prior and the likelihood will regulate the posterior, allowing it to take a standard form. As a result, it can be used to simplify sample collection in discrete mixture models, and also forms the foundation for using the missing data model (Congdon, 2005).

The concept of data augmentation comes from missing value problems, as demonstrated by its similarity to missing cells in balanced two-way tables. From a Bayesian perspective, the posterior distribution of the parameter of interest should be calculated. This can be accomplished through data augmentation, maximising the likelihood estimate, and computing the posterior distribution. The posterior distribution of the parameters of interest must be computed at that very moment. In the event that one is able to use data augmentation in computing for the maximum possible estimate, it is also then possible to use the same data in computing the posterior distribution. Accomplishing this will require the use of the observed data z, which has to be augmented by the unobserved data, or the quantity y. The assumption is that the computation of the augmented data posterior which \( p(\theta|z,y) \) represents can be computed if the values of z and y are known.

However, when figuring the posterior density, the equation \( p(\theta|z) \) must be used. It would be more effective, however, if one can generate multiple values of \( y \) from the predictive distribution \( p(y|z) \) (i.e., multiple imputations of \( y \)). Then \( p(\theta|z) \) can be approximately obtained as the average of \( p(\theta|z,y) \) over the imputed
2’s. However, \( p(y|z) \) must, thereby, turn on \( p(\theta|z) \). So, \( p(\theta|z) \) if it is known, in contrast, then it could be used to calculate \( p(y|z) \). Analytically speaking, this analysis is essentially the method of successive substitution for solving an operator fixed point equation. This fact is routinely exploited, and yields fact that it is this that proves convergence, under mild regularity conditions (Tanner and Wong, 1987).

### 3.12 Model Comparison

The DIC is a goodness of fit or model comparison statistic that takes into account the number of unknown parameters in the model (see Spiegelhalter et al. (2002)). This statistic is intended as a generalisation of the Akaike Information Criterion (AIC; Akaike (1973)). Under a competing model \( M_k \) with a vector of unknown parameters \( \theta_k \) of dimension \( d_k \), let \( \{\theta_k^{(t)}: t = 1,\ldots,T\} \) be a sample of observations simulated from the posterior distribution. The DIC for \( M_k \) is computed as follows:

\[
DIC_k = -\frac{2}{T} \sum_{t=1}^{T} \log p(Z | \theta_k^{(t)}; M_k) + 2d_k
\]  

(3.29)

where

\[ k = 1,\ldots,p, \]

\( Z \): observed ordered categorical or dichotomous data,

\( d_k \): dimension of parameters.

The model having the smallest value of DIC is selected in model comparison (Lee, 2007). When applying DIC practically, it is worth noting that in case the variation in the DIC is minute, and the inferences being made by the models are very different, reporting the model having the smallest DIC might be misleading. DIC may also be applied to non-tested models. As with the AIC and the Bayes factor (Kass and Raftery, 1995), DIC ensures a clear conclusion for supporting the alternative or null hypothesis.
To illustrate the use of the DIC for model comparison, we analysed the same data from a non-linear structural equation model with the same measurement model. Due to the complex model with non-linear fixed covariate and latent variables, and WinBUGS program doesn’t treat with non-linear models so, the DIC is grey out in WinBUGS. We developed a WinBUGS program to treat with non-linear models and find the DIC values of the model. The DIC values corresponding to the non-linear structural equation models with simulation data and a case study are produced by OpenBUGS program.

When making comparisons between complex hierarchical models, there is normally the challenge of some parameters not being defined clearly. With the methods development Markov chain Monte Carlo (MCMC), an investigator would be able to fit considerably huge classes of models. Naturally, this ability leads to the wish of comparing alternative model formulations. This would enable the identification of a class of concise models that seem to elaborate the data’s information sufficiently. For instance, an investigator might consider the need of incorporating a random effect, which would permit overdispersion. Model comparison, as far as the classical modeling structure is concerned, takes place through the measure of fit definition. It is all about the deviance statistic (Celeux et al., 2006).
Figure 3.6: Flow Chart of the Proposed Methods
REFERENCES


0022-0981.


