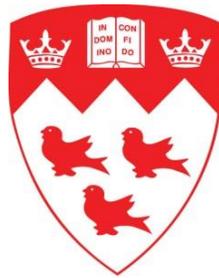


Bayesian Generalized Structured Component Analysis

Ji Yeh Choi

Doctor of Philosophy



Department of Psychology
McGill University, Montreal, Canada
March 2017

A dissertation submitted to McGill University in partial fulfillment of
the requirements of the degree of Doctor of Philosophy.

Copyright © 2017 Ji Yeh Choi

Preface and Statement of Original Contribution

The present research develops a new method that integrates GSCA to a Bayesian framework. The method provides several appealing features compared to original GSCA. It allows fitting various structural equation models while accounting for measurement errors in the observed variables. In addition, the method provides additional fit measures for model assessment and model comparison from the Bayesian perspectives. I, Ji Yeh Choi, am the primary author of this dissertation and have conceptualized and written this dissertation in its entirety. My doctoral supervisor Professor Heungsun Hwang played an essential role in the supervision of my dissertation research, providing important feedback and guidance at every stage.

Acknowledgements

Undertaking this Ph.D. degree has been a life-changing experience for me, and truly I could not have been accomplished this without the support and guidance that I received from many people. Here, I would like to take this opportunity to express my heartfelt gratitude and sincere appreciation to them.

Foremost, I am forever indebted to my supervisor Professor Heungsun Hwang for tremendous support and encouragements he gave me during graduate years at McGill University. As my most admired mentor, his extensive expertise and insights taught me compendious knowledge required to be an erudite quantitative psychologist. I have been very privileged to have him as my supervisor whose generous advice, inspiring guidance, and considerate understanding enabled me to develop and complete this dissertation. I also want to thank his family the hearty welcomes and the kindness for inviting me over to house parties around Christmas and Thanksgiving.

I am grateful to a number of faculty members in the Department of Psychology at McGill University. In particular, I like to express my gratitude to Professor Yoshio Takane who had guided me as my advisory committee members in my early graduate years. During quantitative lab meetings held bi-weekly, he had provided instrumental advices and suggestions for various challenges I encountered while undergoing research. I would like to thank the other members of my committee as well: Professor Hsiu-Ting Yu for always giving me constructive suggestions and insights for the reserach, and Professor Fei Gu for bringing new perspectives on various issues in Psychometrics. I extend my appreciation to Professor Amsel Rhonda for her hospitality and continual support. I like to thank all staffs in the department for their help, especially Giovanna LoCascio.

I am pleased to have a number of former and current lab members in our quantitative psychology program: Kwanghee Jung, Hye Won Suk, Lixing Zhou, Tianyu Tan, Sunmee Kim, Ramsey Cardwell, Newman Knowlton, and Jungkyu Park. I appreciate unforgettable memories with them that I could not share elsewhere. I look forward to having all as collaborators in the years ahead.

I express my deepest gratitude to my family, my parents, my brother, and my husband, to whom this work is dedicated, for their unconditional love and unwavering support. I won't be this stronger without them; they are the reason why I keep moving forward. Special thanks to my husband Jaeho for being there for me.

Finally, I would like to thank God for giving me the power to believe in my passion and pursue my dreams, and being there every step of the way through this journey. In Your strong hands, I place my life today, choosing to depend of You to light and guide my way.

Abstract

Generalized structured component analysis (GSCA) is a component-based approach to structural equation modeling (SEM) that postulates and examines various directional relationships among latent and observed variables. GSCA constructs components or weighted composites of observed variables as proxies for latent variables. It combines three sub-models, such as measurement, structural, and weighted relation models, into a unified formulation, and estimates all model parameters simultaneously via least squares. Over the past decade, GSCA has been extended to deal with a wider range of data types including discrete, multilevel, or intensive longitudinal data, as well as to accommodate a more variety of complex analyses such as latent moderation analysis, the capturing of cluster-level heterogeneity, and regularized analysis. To date, nonetheless, there has been no attempt to generalize the scope of GSCA into the Bayesian framework. In this dissertation, a novel extension of GSCA, called Bayesian GSCA, is proposed that estimates parameters within the Bayesian framework. Bayesian GSCA can be more attractive than GSCA in numerous respects. Firstly, it infers the probability distributions of parameters, treating the parameters as random variables, which in turn facilitates the interpretation of the parameters. Secondly, it permits specifying various structures of error terms in the measurement model, which are left unspecified in GSCA. Thirdly, it provides additional fit measures for model assessment and comparison from the Bayesian perspectives. Lastly, it allows directly incorporating external information on parameters, which may be obtainable from past research, expert opinions, subjective beliefs or knowledge on the parameters, as the form of prior distributions in the modelling process. Bayesian GSCA adopts a Markov Chain Monte Carlo method, i.e., Gibbs Sampler, to update the posterior distributions for parameters. The dissertation begins by describing two building blocks of Bayesian GSCA – GSCA and Bayesian

inference, and subsequently discusses the technical underpinnings of Bayesian GSCA. It also demonstrates the usefulness of Bayesian GSCA based on the analyses of both simulated and real data.

Abrégé

L'analyse en composantes structurée généralisée (ACSG) est une approche de la modélisation par équations structurelles (MES) basée sur les composantes qui postule et examine divers liens directionnels entre des variables latentes et observées. L'ACSG construit des composantes ou des ensembles pondérés de variables observées pour représenter des variables latentes. Elle combine trois sous-modèles, soit le modèle de mesure, le modèle structurel et le modèle des liens pondérés dans une formulation unifiée et estime tous les paramètres des modèles simultanément via la méthode des moindres carrés. Durant les dix dernières années, l'ACSG a évolué pour traiter une plus grande gamme de types de données, notamment des données discrètes, multiniveaux, et longitudinales avec de nombreuses mesures sur une longue durée, ainsi que pour s'adapter à une plus grande variété d'analyses complexes telles que l'analyse de modération avec des variables latentes, la mesure de l'hétérogénéité au niveau d'un sous-groupe et l'analyse régularisée. À date, néanmoins, il n'y a pas encore eu de tentative de généralisation de l'ACSG dans le cadre bayésien. Dans cette thèse, une nouvelle extension de l'ACSG, appelée ACSG bayésienne, évaluant les paramètres à l'intérieur du cadre bayésien, est proposée. L'ACSG bayésienne peut se prouver plus attractive que l'ACSG à de nombreux égards. Premièrement, elle déduit la loi de probabilité des paramètres, en traitant les paramètres comme des variables aléatoires, ce qui facilite alors l'interprétation des paramètres. Deuxièmement, elle permet de spécifier plusieurs structures de terme d'erreur dans le modèle de mesure qui ne sont pas spécifiées dans l'ACSG. Troisièmement, elle fournit des mesures d'ajustement supplémentaires pour l'évaluation et la comparaison du modèle d'une perspective bayésienne. Enfin, elle permet d'incorporer de l'information externe sur les paramètres, qui peut provenir de recherches passées, d'opinions d'experts, de croyances subjectives ou de connaissances sur les

paramètres, directement sous la forme de distributions préalables dans le processus de modélisation. L'ACSG bayésienne adopte une méthode de Monte-Carlo par chaînes de Markov, soit l'échantillonnage de Gibbs, pour mettre à jour les distributions postérieures des paramètres. La thèse commence par une description des deux fondations de l'ACSG bayésienne – l'ACSG et l'inférence bayésienne, avant d'examiner les fondements techniques de l'ACSG bayésienne. Elle démontre également l'utilité de l'ACSG bayésienne pour l'analyse de données simulées et réelles.

Table of Contents

Acknowledgements.....	i
Abstract.....	iii
Abrégé.....	v
Table of Contents.....	vii
List of Tables.....	ix
List of Figures.....	x
Chapter 1. Introduction.....	1
Chapter 2. Generalized Structured Component Analysis.....	8
2.1. Model Specification.....	8
2.2. Parameter Estimation.....	11
Chapter 3. An Overview of Bayesian Inference.....	16
3.1. Two Philosophies for Statistical Inference.....	16
3.2. Bayes' Theorem.....	19
3.3. Markov Chain Monte Carlo (MCMC) Methods.....	20
3.4. Model Assessment and Comparison in Bayesian Inference.....	29
Chapter 4. Bayesian Generalized Structured Component Analysis.....	34
4.1. Model Specification.....	34
4.2. Parameter Estimation.....	38
4.3. Other Extensions.....	48
4.3.1. Correlated error terms in measurement models.....	48
4.3.2. Analysis of categorical data.....	50
Chapter 5. Simulation Study.....	53
5.1. Simulation Study 1.....	53
5.1.1. Simulation Design and Data Generation.....	53
5.1.2. Results.....	55
5.2. Simulation Study 2.....	58

5.2.1. Simulation Design.....	58
5.2.2. Results and Discussion	59
Chapter 6. An Empirical Example	61
6.1. Data Description and Model Specification	61
6.2. Analysis	63
6.3. Results.....	64
Chapter 7. Summary and Discussion	72
References.....	76

List of Tables

Table 3- 1. Differences between Frequentist and Bayesian approaches.....	17
Table 3- 2. A description of the Metropolis-Hastings algorithm.....	23
Table 3- 3. A description of the Gibbs Sampler algorithm.....	25
Table 4- 1. Summary of notations for the variables in BGSCA.	39
Table 4- 2. Specification of BGSCA sub-models and distributional assumptions.	40
Table 5- 1. The average values of mean square error (MSE) of the estimates in the weighted relation model obtained from BGSCA.	56
Table 5- 2. The average values of MSE of the estimates in the measurement and structural model obtained from BGSCA.....	57
Table 5- 3. The average values of MSE of the estimates in the structural model obtained from BGSCA.	58
Table 5- 4. Specification of different hyperparameter values.....	58
Table 5- 5. The average values of MSE of the estimates of weights, loadings, and path coefficients across four different specifications of hyperparameter values.....	60
Table 6- 1. Twenty-one observed variables for four latent variables in Bergami and Bagozzi's (2000) organizational identification data.	62
Table 6- 2. Results of the weight estimates obtained from BGSCA.	67
Table 6- 3. Results of the loading estimates obtained from BGSCA.	68
Table 6- 4. Results of the path coefficient estimates obtained from BGSCA.	70
Table 6- 5. Different Bayes factors and their interpretations.....	71

List of Figures

Figure 2- 1. A hypothetical example of a structural equation model.....	10
Figure 4- 1. A hypothetical BGSCA model.....	36
Figure 5- 1. The structural model specified for simulation study 1.....	53
Figure 6- 1. A hypothesized model for the organizational identification data	63
Figure 6- 2. Trace plots of the estimates of (a) weights, (b) loadings, and (c) path coefficients..	65
Figure 6- 3. ACF plots of the estimates of (a) weights, (b) loadings, and (c) path coefficients...	66
Figure 6- 4. A hypothetical model specified for model comparison.	70

Chapter 1. Introduction

Structural Equation Modeling (SEM) is used for examining various directional relationships among observed variables and hypothetical constructs, also called latent variables, which are not directly measurable. SEM is comprised of two sub-models: measurement and structural models. The measurement model specifies the hypothesized relationships between latent and observed variables, whereas the structural model reflects the hypothesized directional relationships among latent variables.

In general, there are two different domains of SEM: factor-based (Jöreskog, 1969) and component-based (Wold, 1975; Hwang & Takane, 2004). As the names suggest, the former was developed in the framework of common factor analysis, whereas the latter originated from component analysis. Specifically, factor-based SEM assumes that a latent variable can be approximated by a factor accounting only for the common variance of an observed variable. Since basic factor-based structural equation models were formulated as a combination of confirmatory factor analytic and path analytic models for continuous data (e.g., Jöreskog, 1969), they have been extended and elaborated over the past several decades by relaxing various assumptions, such as multivariate normality of observed variables or linear models, or by taking into account more complex data types. A few examples of such extensions include non-linear SEM accommodating interaction or quadratic terms of latent variables (e.g., Jöreskog & Yang, 1996; Kenny & Judd, 1984; Klein & Moosbrugger, 2000; Schumacker & Marcoulides, 1998), multilevel SEM for handling nested structures of data (e.g., Bollen & Curran, 2006; Muthén, 1994; Satorra & Muthén, 1995), and SEM for non-normal or discrete data (e.g., Chou, Bentler, & Satorra, 1991; Finney & DiStefano, 2006; Hau & Marsh, 2004; Moustaki, 2003; Muthén, 1984; Shimizu & Kano, 2008).

On the other hand, component-based SEM combines component analytic and path analytic models. In component-based SEM, a weighted composite or component of a set of observed variables is assumed to be a proxy for a latent variable, capturing the most representative variation of each of the observed variables. The first approach to component-based SEM is partial least squares path modeling (PLS-PM), which was initiated by Wold (1975, 1982) and became popularized with a number of user-friendly software programs being available, including PLS Graph (Chin, 2001), LVPLS (Lohmöller, 1984), and SmartPLS (Ringle, Wende, & Will, 2005). PLS-PM has some practical advantages over factor-based SEM. For example, it does not require any distributional assumption and is less likely to encounter non-convergence or convergence to improper solutions even in small samples. Nonetheless, PLS-PM carries out two separate stages sequentially to estimate parameters, i.e., one stage for estimating components and the other for estimating remaining parameters such as path coefficients and loadings (e.g., Hwang, Takane, & Tenenhaus, 2015; Tenenhaus, Vinzi, Chatelin, & Lauro, 2005). In the first stage, an iterative algorithm (Wold, 1982, 1985; Lohmöller, 1989) is used to estimate the components and their weights, although it is unknown what optimization criterion this algorithm generally seeks to minimize or maximize. In the second stage, the remaining parameters are estimated by applying a series of linear regression analyses based on the component estimates from the first stage. In other words, PLS-PM has no global optimization criterion for estimating all the parameters in both stages simultaneously. This makes it difficult to calculate overall fit measures that are often proportional to the value of the global optimization criterion at convergence (McDonald, 1996).

More recently, generalized structured component analysis (GSCA; Hwang & Takane, 2004, 2014) was introduced as another component-based approach to SEM. In GSCA, another

sub-model, called the weighted relation model, is specified to explicitly express a latent variable as a component of a set of observed variables. GSCA combines the three sub-models, i.e., measurement, structural, and weighted relation models, into a single formulation. It aims to minimize a global least squares optimization criterion, which is derived from the single formulation, to estimate all parameters simultaneously. Accordingly, unlike PLS-PM, GSCA allows the calculation of overall fit measures to evaluate how well a given model fits to the data as a whole and to compare competing models. In addition, GSCA can deal with a wider range of data types, including multi-level data (Hwang, Takane, & Malhotra, 2007), time-dependent repeated measures data (Jung, Takane, Hwang, & Woodward, 2012), categorical data (Hwang & Takane, 2010), and functional data (Suk & Hwang, 2016). Furthermore, it can conduct a more variety of complex analyses in a technically more coherent and straightforward manner, including multiple group analysis with the imposition of cross-group equality constraints (Hwang and Takane, 2004), the analysis of interaction terms of latent variables (Hwang, Ho, & Lee, 2010), and regularized analysis (Hwang, 2009). For these advantages, GSCA can be considered an alternative to PLS-PM. Thus, this dissertation focuses on GSCA as the main approach to component-based SEM. Free online software (available at <http://www.sem-gesca.org/>) and an R package, `gesca` (Hwang, Kim, Lee, & Park, 2016) are currently available for implementing various GSCA models.

Owing to the revolution of computer-intensive sampling methods, Bayesian approaches to SEM have grown in popularity, particularly for factor-based SEM, including those for standard linear SEM (Muthén & Asparouhov, 2012; Scheines, Hoijtink, & Boomsma, 1999), nonlinear SEM (Arminger & Muthén, 1998; Lee, 2006; Lee, 2007; Lee & Zhu, 2000), and multilevel SEM (Ansari, Jedidi, & Jagpal, 2000; Song & Lee, 2004; Lee & Song, 2012). Lee

(2012) provides a comprehensive review of Bayesian extensions of factor-based SEM. In general, there are three advantages of adopting a Bayesian approach to factor-based SEM, although they are not just unique to this methodology, but rather applicable to other methodological extensions of Bayesian inference. Firstly, a Bayesian approach allows for probabilistic interpretations of parameters by treating the parameters as random variables. By being random, a variable can take on a set of values with probability and thus should be described by a probability distribution. In turn, this provides simpler interpretations on probability values (i.e., p -values) for testing statistical hypotheses as well as interval estimates of parameters, as compared to a non-Bayesian approach (e.g., Congdon, 2007; Kaplan & Depaoli, 2012; Song & Lee, 2012). For example, in the non-Bayesian approach, a 95 % confidence interval of a parameter estimate is a range of values that 95% of the intervals computed from repeated sampling would contain the parameter. This suggests that the probability of containing the parameter value in a particular interval is either zero or one. On the other hand, in the Bayesian approach, a 95 % confidence interval estimate (also called credible interval; Edwards, Lindman & Savage, 1963) is directly interpreted as an interval that includes the parameter value with a 95 % probability. Similarly, the interpretation of a p -value is more intuitively appealing in the Bayesian approach because it simply indicates the posterior probability of the null hypothesis given the data, whereas the non-Bayesian approach's p -value indicates the probability of obtaining a test statistic value at least as extreme as the one observed given that the null hypothesis is true. Hence, a Bayesian approach provides more intuitive interpretations of parameters with probabilistic statements by deriving their posterior distributions.

Another notable advantage of adopting a Bayesian approach is that it can incorporate external information on parameters, such as previous research findings or expert opinions, which

might otherwise be neglected in the modeling process. A Bayesian approach enables to assign probability distributions to the parameters, called prior distributions, by formally quantifying uncertainty about the parameters (e.g., Hoff, 2009; Gelman, Carlin, Stern, & Rubin, 2004). This allows integrating any relevant information about the parameters into the modeling process and enhances the accuracy of parameter estimates, unless the specified prior information is completely incorrect.

Lastly, a Bayesian approach would perform well even when sample size is small (e.g., Muthén & Asparouhov, 2012; Palomo, Dunson, Bollen, 2007; Scheines et al., 1999). This is because the Bayesian approach does not rely on asymptotic theory (i.e., large-sample approximations), but rather it simulates samples from a posterior distribution and computes test statistics by summarizing the posterior distribution itself.

In contrast to the Bayesian developments in factor-based SEM, to date, there has been no attempt to apply a Bayesian approach to GSCA. It is partially attributed to the fact that parameters in GSCA are estimated and tested without imposing any distributional assumptions. This distribution-free feature had provided little rationale to apply to GSCA an alternative estimation method that in fact requires some extent of distributional assumptions. However, GSCA can benefit from adopting a Bayesian approach in several respects. In addition to the advantages of Bayesian factor-based SEM delineated above, a Bayesian extension of GSCA refines the original GSCA by providing three more features in the modeling process. Firstly, it allows estimating the true measurements of observed data, eliminating random errors. The updated true measurements are then used in place of the observed data for specifying the sub-models in GSCA. This would be preferable to using original observed data themselves, particularly when the observed data are greatly contaminated by a large magnitude of random

errors. As the prior distributions for unknown parameters are specified, one may also formulate a prior distribution for the true measurements, integrating any relevant background knowledge, which is possibly available from previous research (e.g., mean values or dispersions of the true measurements), so as to improve the accuracy of their estimates. Secondly, unlike the original GSCA, the Bayesian extension of GSCA permits specifying parameters associated with measurement error terms (e.g., the means, variances, and covariances of the error terms) as an additional set of parameters to be estimated. Thirdly, a Bayesian extension of GSCA can offer more flexibility in assessing the adequacy of a hypothesized model than what the original GSCA currently offers. Technically, model fit measures that are commonly used in factor-based SEM and/or their corresponding Bayesian approaches can also be obtained in the Bayesian extension of GSCA. As a special case, when uniform prior distributions are assigned to all parameters, the goodness of fit of the GSCA model can be statistically tested using a chi-square statistic, as in factor-based SEM. One may also utilize a range of other fit criteria to test overall model fit, for example, GFI, SRMR, RMSEA, NFI, and CFI (e.g., Mulaik, 2009, Chapter 15), which are available for factor-based SEM. When there exist several theoretically plausible alternative models, the Bayesian extension of GSCA further enables to compare them, using the Bayes factor (Kass & Raftery, 1995) or its approximated information criteria, BIC (Kass & Raftery, 1995) and DIC (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002).

In this dissertation, I thus propose *Bayesian generalized structured component analysis* (BGSCA), which integrates the original GSCA (Hwang & Takane, 2004) into a Bayesian framework. More specifically, in the weighted relation model of BGSCA, latent variables are defined as weighted composites of so-called error-free data that are considered true measurements of observed data, eliminating noises or random errors. The posterior distributions

for the parameters in the three sub-models of BGSCA are then estimated via the Gibbs Sampler algorithm (Geman & Geman, 1984).

The remaining chapters of the dissertation are organized as follows. Chapters 2 and 3 provide descriptions of the two building blocks of BGSCA – GSCA and Bayesian inference. Chapter 4 discusses the technical underpinnings of BGSCA. Chapter 5 reports two simulation studies that examine the accuracy of parameter recovery of BGSCA and the sensitivity of BGSCA results against different specifications of prior distributions. Chapter 6 illustrates the empirical usefulness of BGSCA by applying it to a real data set. The final chapter summarizes the implications of BGSCA and discusses potential topics for future research.

Chapter 2. Generalized Structured Component Analysis

As stated in the previous chapter, GSCA is proposed for component-based SEM that investigates a variety of directional relationships among latent variables as well as the relationships between latent and observed variables, in which latent variables are defined as weighted composites of observed variables. In this chapter, the original GSCA model and its parameter estimation are recapitulated in order to facilitate an understanding of Bayesian GSCA that shall be discussed in Chapter 4.

2.1. Model Specification

GSCA specifies three sub-models to construct a general structural equation model. The three sub-models are measurement, structural, and weighted relation models (e.g., Hwang & Takane, 2014, Chapter 2). Let \mathbf{x}_i denote a J_1 by 1 vector of observed exogenous variables for the i th subject ($i = 1, \dots, I$). Let \mathbf{y}_i denote a J_2 by 1 vector of observed endogenous variables for the i th subject. The measurement model specifies the relationships between observed variables and their latent variables, as follows

$$\begin{aligned}\mathbf{x}_i &= \mathbf{C}^{(x)'} \boldsymbol{\gamma}_i^{(x)} + \boldsymbol{\delta}_i^{(x)} \\ \mathbf{y}_i &= \mathbf{C}^{(y)'} \boldsymbol{\gamma}_i^{(y)} + \boldsymbol{\delta}_i^{(y)},\end{aligned}\tag{2.1}$$

where $\boldsymbol{\gamma}_i^{(x)}$ and $\boldsymbol{\gamma}_i^{(y)}$ are P_1 by 1 and P_2 by 1 vectors of the latent variable scores for the i th subject, $\mathbf{C}^{(x)}$ and $\mathbf{C}^{(y)}$ are P_1 by J_1 and P_2 by J_2 matrices of loadings relating the latent variables to the corresponding observed variables, respectively. In (2.1), $\boldsymbol{\delta}_i^{(x)}$ and $\boldsymbol{\delta}_i^{(y)}$ denote J_1 by 1 and J_2 by 1 vectors of measurement errors for the observed variables for the i th subject. Let $\mathbf{z}_i = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{bmatrix}$ and

$\boldsymbol{\Gamma}_i = \begin{bmatrix} \boldsymbol{\gamma}_i^{(x)} \\ \boldsymbol{\gamma}_i^{(y)} \end{bmatrix}$. Then, the two equations in the measurement model can be unified as

$$\mathbf{z}_i = \mathbf{C}'\boldsymbol{\Gamma}_i + \boldsymbol{\delta}_i, \quad (2.2)$$

where $\mathbf{C}' = \begin{bmatrix} \mathbf{C}^{(x)'} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{(y)'} \end{bmatrix}$ and $\boldsymbol{\delta}_i = \begin{bmatrix} \boldsymbol{\delta}_i^{(x)} \\ \boldsymbol{\delta}_i^{(y)} \end{bmatrix}$.

The structural model defines the relationships between latent variables, expressed as

$$\boldsymbol{\Gamma}_i = \mathbf{B}\boldsymbol{\Gamma}_i + \boldsymbol{\tau}_i, \quad (2.3)$$

where \mathbf{B} is a (P_1+P_2) by (P_1+P_2) matrix of path coefficients reflecting directional relationships among latent variables, and $\boldsymbol{\tau}_i$ is a (P_1+P_2) by 1 vector of errors of latent variables.

As in component analysis, GSCA constructs components or weighted composites of observed variables as proxies for latent variables. The weighted relation model is used to express such relationships between latent and observed variables. This sub-model is written as

$$\boldsymbol{\Gamma}_i = \mathbf{W}'\mathbf{z}_i, \quad (2.4)$$

where $\mathbf{W} = \begin{bmatrix} \mathbf{W}^{(x)} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{(y)} \end{bmatrix}$ is a (J_1+J_2) by (P_1+P_2) matrix of (component) weights assigned to observed variables: in specific, $\mathbf{W}^{(x)}$ is a J_1 by P_1 matrix of weights for the exogenous observed variables, and $\mathbf{W}^{(y)}$ is a J_2 by P_2 for the endogenous observed variables.

For illustration, a hypothetical example of a structural equation model with three latent variables ($P_1+P_2 = 3$) is displayed in Figure 2- 1 . In the figure, two latent variables are considered exogenous ($P_1 = 2$) and the remaining one is endogenous ($P_2 = 1$). The first and second exogenous latent variables are associated with three and two observed variables, respectively, yielding $J_1 = 5$. The endogenous latent variable is associated three observed variables, i.e., $J_2 = 3$. A straight line represents a weight assigned to each observed variable, whereas an arrow signifies either a loading for an observed variable or a path coefficient relating latent variables.

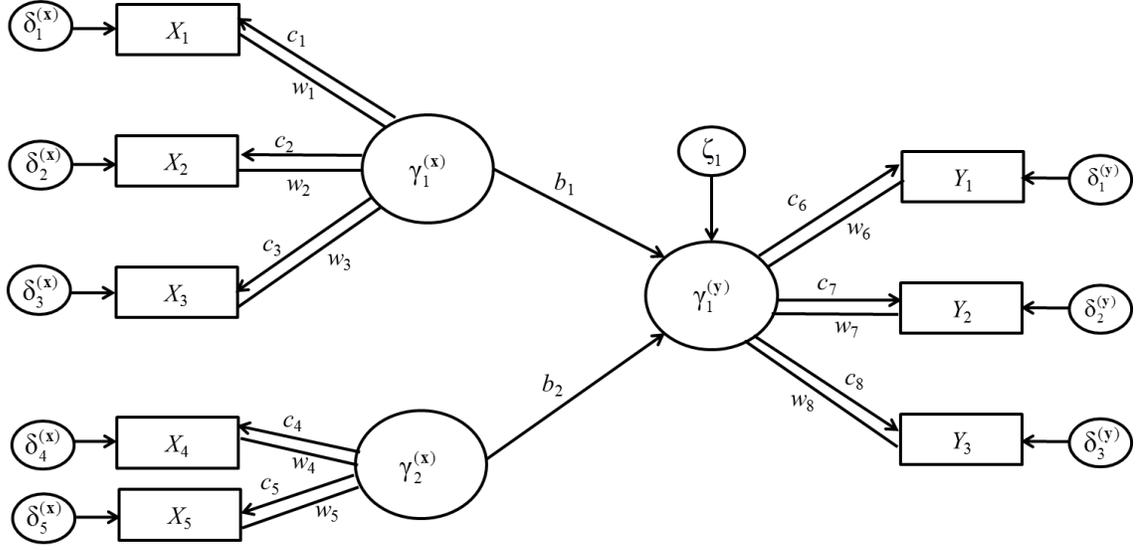


Figure 2- 1. A hypothetical example of a structural equation model.

The measurement model for the hypothetical example is given as

$$\begin{bmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \\ X_{i4} \\ X_{i5} \\ Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{bmatrix} = \begin{bmatrix} c_1 & 0 & 0 \\ c_2 & 0 & 0 \\ c_3 & 0 & 0 \\ 0 & c_4 & 0 \\ 0 & c_5 & 0 \\ 0 & 0 & c_6 \\ 0 & 0 & c_7 \\ 0 & 0 & c_8 \end{bmatrix} \begin{bmatrix} \gamma_{i1}^{(x)} \\ \gamma_{i2}^{(x)} \\ \gamma_{i1}^{(y)} \end{bmatrix} + \begin{bmatrix} \delta_{i1}^{(x)} \\ \delta_{i2}^{(x)} \\ \delta_{i3}^{(x)} \\ \delta_{i4}^{(x)} \\ \delta_{i1}^{(y)} \\ \delta_{i2}^{(y)} \\ \delta_{i3}^{(y)} \end{bmatrix}, \quad (2.5)$$

$$\mathbf{z}_i = \mathbf{C}'\mathbf{\Gamma}_i + \boldsymbol{\delta}_i,$$

where $\mathbf{C}^{(x)} = \begin{bmatrix} c_1 & c_2 & c_3 & 0 & 0 \\ 0 & 0 & 0 & c_4 & c_5 \end{bmatrix}$ and $\mathbf{C}^{(y)} = [c_6 \ c_7 \ c_8]$. The structural model is given as

$$\begin{bmatrix} \gamma_{i1}^{(x)} \\ \gamma_{i2}^{(x)} \\ \gamma_{i1}^{(y)} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ b_1 & b_2 & 0 \end{bmatrix} \begin{bmatrix} \gamma_{i1}^{(x)} \\ \gamma_{i2}^{(x)} \\ \gamma_{i1}^{(y)} \end{bmatrix} + \begin{bmatrix} \gamma_{i1}^{(x)} \\ \gamma_{i2}^{(x)} \\ \zeta_{i1} \end{bmatrix}, \quad (2.6)$$

$$\mathbf{\Gamma}_i = \mathbf{B}\mathbf{\Gamma}_i + \boldsymbol{\tau}_i.$$

The weighted relation model for this hypothetical example is specified as

$$\begin{bmatrix} \gamma_{i1}^{(x)} \\ \gamma_{i2}^{(x)} \\ \gamma_{i1}^{(y)} \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & w_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & w_4 & w_5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_6 & w_7 & w_8 \end{bmatrix} \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \\ x_{i5} \\ y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix}, \quad (2.7)$$

$$\Gamma_i = \mathbf{W}'\mathbf{z}_i,$$

$$\text{where } \mathbf{W}^{(x)} = \begin{bmatrix} w_1 & 0 \\ w_2 & 0 \\ w_3 & 0 \\ 0 & w_4 \\ 0 & w_5 \end{bmatrix} \text{ and } \mathbf{W}^{(y)} = \begin{bmatrix} w_6 \\ w_7 \\ w_8 \end{bmatrix}.$$

2.2. Parameter Estimation

As indicated in the above equations, the GSCA model contains three sets of parameters to be estimated: loadings (\mathbf{C}), path coefficients (\mathbf{B}), and weights (\mathbf{W}). These unknown parameters are estimated by minimizing the following least-squares criterion:

$$\phi = \sum_{i=1}^I \delta_i' \delta_i + \tau_i' \tau_i, \quad (2.8)$$

subject to the standardization constraint $\sum_{i=1}^I \gamma_{ip_1}^{(x)2} = I$ and $\sum_{i=1}^I \gamma_{ip_2}^{(y)2} = I$. This least-squares criterion

is equivalent to the sum of all squared errors in both measurement and structural models in (2.2) and (2.3) over I subjects. An alternating least-squares (ALS) algorithm (de Leeuw, Young, & Takane, 1976) was developed to minimize the criterion (Hwang & Takane, 2004).

The ALS algorithm alternates three main steps until convergence. In the first step, for fixed \mathbf{B} and \mathbf{W} , \mathbf{C} is updated in the least-squares manner. Because \mathbf{C} is involved in the measurement model only, this is equivalent to minimizing the following criterion:

$$\begin{aligned}
 \phi_1 &= \sum_{i=1}^I \boldsymbol{\delta}_i' \boldsymbol{\delta}_i \\
 &= \text{SS}(\mathbf{Z} - \boldsymbol{\Gamma}^* \mathbf{C}) \\
 &= \text{SS}(\text{vec}(\mathbf{Z}) - (\mathbf{I} \otimes \boldsymbol{\Gamma}^*) \text{vec}(\mathbf{C})),
 \end{aligned} \tag{2.9}$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_I]'$ is an I by $(J_1 + J_2)$ matrix of observed variables, $\boldsymbol{\Gamma}^* = [\boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_I]'$ is an I by $(P_1 + P_2)$ matrix of latent variables, $\text{SS}(\mathbf{A}) = \text{tr}(\mathbf{A}'\mathbf{A})$, $\text{vec}(\mathbf{Z})$ is a super-vector obtained by stacking the columns of \mathbf{Z} in order, and \otimes indicates the Kronecker product. Given that \mathbf{C} contains constants such as zeros as shown in (2.5), (2.9) can be rewritten as

$$\phi_1 = \text{SS}(\text{vec}(\mathbf{Z}) - \boldsymbol{\Xi}_1 \mathbf{c}^\oplus), \tag{2.10}$$

where \mathbf{c}^\oplus is a vector formed by eliminating all zero elements from $\text{vec}(\mathbf{C})$, and $\boldsymbol{\Xi}_1$ is a matrix formed by eliminating the columns of $\mathbf{I} \otimes \boldsymbol{\Gamma}^*$ corresponding to the zero elements in $\text{vec}(\mathbf{C})$. By solving $\partial \phi_1 / \partial \mathbf{c}^\oplus = \mathbf{0}$, the least-squares estimate of \mathbf{c}^\oplus is given by

$$\hat{\mathbf{c}}^\oplus = (\boldsymbol{\Xi}_1' \boldsymbol{\Xi}_1)^{-1} \boldsymbol{\Xi}_1' \text{vec}(\mathbf{Z}). \tag{2.11}$$

The updated $\hat{\mathbf{c}}^\oplus$ is reshaped into \mathbf{C} .

In the second step, \mathbf{B} is updated for fixed \mathbf{C} and \mathbf{W} . As \mathbf{B} is involved in the structural model only, this is equivalent to minimizing

$$\begin{aligned}
 \phi_2 &= \sum_{i=1}^I \boldsymbol{\tau}_i' \boldsymbol{\tau}_i \\
 &= \text{SS}(\boldsymbol{\Gamma}^* - \boldsymbol{\Gamma}^* \mathbf{B}) \\
 &= \text{SS}(\text{vec}(\boldsymbol{\Gamma}^*) - (\mathbf{I} \otimes \boldsymbol{\Gamma}^*) \text{vec}(\mathbf{B})) \\
 &= \text{SS}(\text{vec}(\boldsymbol{\Gamma}) - \boldsymbol{\Xi}_2 \mathbf{b}^\oplus),
 \end{aligned} \tag{2.12}$$

where \mathbf{b}^\oplus is a vector formed by eliminating zero elements from $\text{vec}(\mathbf{B})$, and Ξ_2 is a matrix formed by eliminating the columns of $\mathbf{I} \otimes \Gamma^\star$ corresponding to the zero elements in $\text{vec}(\mathbf{B})$. The least-squares estimate of \mathbf{b}^\oplus is updated by

$$\hat{\mathbf{b}}^\oplus = (\Xi_2' \Xi_2)^{-1} \Xi_2' \text{vec}(\Gamma^\star), \quad (2.13)$$

and the updated \mathbf{B} is reconstructed by putting the estimate of $\hat{\mathbf{b}}^\oplus$ into \mathbf{B} .

In the third step, for fixed \mathbf{C} and \mathbf{B} , \mathbf{W} is updated. Note that \mathbf{W} is involved in both terms in (2.8). The criterion can be rewritten as

$$\begin{aligned} \phi &= \sum_{i=1}^I \delta_i' \delta_i + \tau_i' \tau_i \\ &= \text{SS}(\mathbf{Z} - \Gamma^\star \mathbf{C}) + \text{SS}(\Gamma^\star - \Gamma^\star \mathbf{B}) \\ &= \text{SS}(\mathbf{Z} - \mathbf{ZWC}) + \text{SS}(\mathbf{ZW} - \mathbf{ZWB}) \\ &= \text{SS}(\mathbf{Z} [\mathbf{I} \ \mathbf{W}] - \mathbf{ZW} [\mathbf{C} \ \mathbf{B}]) \\ &= \text{SS}(\mathbf{ZT} - \mathbf{ZWA}), \end{aligned} \quad (2.14)$$

where \mathbf{I} is the identity matrix of size $J_1 + J_2$, $\mathbf{T} = [\mathbf{I} \ \mathbf{W}]$ is a $(J_1 + J_2)$ by $(J_1 + J_2 + P_1 + P_2)$ matrix, and $\mathbf{A} = [\mathbf{C} \ \mathbf{B}]$ is a $(P_1 + P_2)$ by $(J_1 + J_2 + P_1 + P_2)$ matrix. Because only some columns of \mathbf{W} are duplicated in both \mathbf{T} and \mathbf{W} in (2.14), each column of \mathbf{W} is separately updated at a time. Let \mathbf{w}_p denote the p th column of \mathbf{W} ($p = 1, \dots, (P_1 + P_2)$). Let $\mathbf{T}_{(-p)}$ denote \mathbf{T} whose p th column is replaced by a vector of zeros, and $\mathbf{T}_{(p)}$ denote \mathbf{T} whose columns are all zero vectors except the p th column (i.e., $\mathbf{T} = \mathbf{T}_{(-p)} + \mathbf{T}_{(p)}$). Let $\mathbf{A}_{(-p)}$ be a product matrix of \mathbf{W} whose p th column is the vector of zero and \mathbf{A} . Then, to update \mathbf{w}_p , (2.14) can be rewritten as:

$$\begin{aligned}
 \phi_3 &= \text{SS}\left(\left(\mathbf{Z}\mathbf{T}_{(-p)} + \mathbf{Z}\mathbf{T}_{(p)}\right) - \left(\mathbf{Z}\mathbf{\Lambda}_{(-p)} + \mathbf{Z}\mathbf{w}_p \mathbf{a}_p\right)\right) \\
 &= \text{SS}\left(\left(\mathbf{Z}\mathbf{T}_{(-p)} - \mathbf{Z}\mathbf{\Lambda}_{(-p)}\right) - \left(\mathbf{Z}\mathbf{w}_p \mathbf{a}^* - \mathbf{Z}\mathbf{T}_{(p)}\right)\right) \\
 &= \text{SS}\left(\mathbf{\Lambda} - \left(\mathbf{Z}\mathbf{w}_p \mathbf{a}_p - \mathbf{Z}\mathbf{w}_p \mathbf{u}_p\right)\right) \\
 &= \text{SS}\left(\mathbf{\Lambda} - \left(\mathbf{Z}\mathbf{w}_p \left(\mathbf{a}_p - \mathbf{u}_p\right)\right)\right) \\
 &= \text{SS}\left(\text{vec}(\mathbf{\Lambda}) - \left(\left(\mathbf{a}_p - \mathbf{u}_p\right)' \otimes \mathbf{Z}\right) \mathbf{w}_p\right),
 \end{aligned} \tag{2.15}$$

where \mathbf{a}_p is the p th row of \mathbf{A} , and \mathbf{u}_p indicates a 1 by $(J_1+J_2+P_1+P_2)$ vector whose p th element is unity but all the other elements are zeros. Let \mathbf{w}_p^\oplus denote a vector formed by eliminating zero elements in \mathbf{w}_p and $\mathbf{\Xi}_3$ a matrix formed by eliminating the columns of $\left(\mathbf{a}_p - \mathbf{u}_p\right)' \otimes \mathbf{Z}$ corresponding to the zero elements in \mathbf{w}_p . The least-squares estimate of \mathbf{w}_p^\oplus is obtained by

$$\hat{\mathbf{w}}_p^\oplus = \left(\mathbf{\Xi}_3' \mathbf{\Xi}_3\right)^{-1} \mathbf{\Xi}_3' \text{vec}(\mathbf{\Lambda}). \tag{2.16}$$

After reconstructing $\hat{\mathbf{W}}_p$ from $\hat{\mathbf{w}}_p^\oplus$, the updated $\hat{\mathbf{W}}_p$ is multiplied by $\sqrt{\frac{I}{\mathbf{w}_p' \mathbf{Z}' \mathbf{Z} \mathbf{w}_p}}$ to satisfy the

standardization constraint $\sum_{i=1}^I \gamma_{i p_1}^{(x)^2} = I$ and $\sum_{i=1}^I \gamma_{i p_2}^{(y)^2} = I$. The abovementioned three steps can

reduce to two steps by updating both \mathbf{C} and \mathbf{B} in a single step (see Hwang and Takane, 2004).

However, to facilitate an understanding of the estimation algorithm that shall be proposed for

Bayesian GSCA in Chapter 4, separate steps for estimating \mathbf{C} and \mathbf{B} are presented in this chapter.

Also refer to Chapter 2 in Hwang and Takane (2014) or the Appendix in Hwang, DeSarbo, and

Takane (2007) for more details on the two-step ALS algorithm.

In GSCA, an overall fit measure, called FIT (Hwang & Takane, 2004) can be calculated to assess whether a given hypothesized model fits well to the data. The FIT indicates the total

variance of all observed and latent variables accounted for by a specified model, which is given by:

$$\text{FIT} = 1 - \frac{\phi}{\text{SS}(\mathbf{ZT})}, \quad (2.17)$$

where $\text{SS}(\mathbf{ZT})$ is the sum of squares of all observed and latent variables in (2.14). The values of FIT range from 0 to 1. The larger FIT value, the more variance is explained. The Adjusted FIT (AFIT) is also available as a variant of FIT, which takes into account model complexity (Hwang, DeSarbo, & Takane, 2007). Moreover, there are two additional measures of overall model fit, Goodness-of-fit index (GFI) and standardized root mean square residual (SRMR). These measures evaluate the discrepancy between the sample covariances and model-implied covariances re-reproduced by model parameter estimates. More details on these overall fit measures can be found in Hwang and Takane (2014, Chapter 2).

Chapter 3. An Overview of Bayesian Inference

This chapter provides a brief introduction to Bayesian inference. In Section 3.1, main characteristics distinguishing a Bayesian approach from a non-Bayesian approach, also called classical or Frequentist approach, are discussed, followed by a description of key sources of information required for adopting a Bayesian approach. In Section 3.2, the process of deriving posterior distributions for unknown parameters via Bayes' theorem is explained. Bayes' theorem is a basis for obtaining posterior probabilities of parameters given data and updating one's existing beliefs in the light of new evidence collected from the data. In Section 3.3, computer-intensive sampling methods, referred to as Markov Chain Monte Carlo (MCMC) methods, are described for updating the posterior distributions and conducting Bayesian inference. This section also provides an example of applying a Bayesian approach to a simple linear regression analysis through the implementation of an MCMC method, called Gibbs Sampler. In Section 3.4, model evaluation and comparison in Bayesian inference are presented.

3.1. Two Philosophies for Statistical Inference

When making statistical inference about an unknown parameter, there are two distinct perspectives: Frequentist and Bayesian approaches. The Frequentist approach associates a probability with the relative frequency of outcomes in repeated samples. In this approach, data is considered random while a parameter is unknown but fixed. Accordingly, inference about the parameter is made from a hypothetical distribution of estimates of the parameter generated under repeated sampling of data. On the other hand, the Bayesian approach considers the unknown parameter random and quantifies the uncertainty about the parameter with a probability distribution representing the degree of belief about a certain value of the parameter. Treating an

unknown parameter as a random variable is the foundation of the Bayesian approach. Table 3- 1 summarizes the differences between Frequentist and Bayesian approaches.

Table 3- 1. Differences between Frequentist and Bayesian approaches.

	Frequentist	Bayesian
Nature of parameters in a model	Unknown but fixed; population parameter has a true fixed value.	Unknown and random; population parameter follows a probability distribution.
Definition of probability	Relative frequency.	Degree of belief about the value of a parameter.
Inference	Using sampling distribution of an estimate.	Using the posterior probability distribution for a parameter.

In a Bayesian approach, a probabilistic statement about the parameter is formally made by synthesizing two sources of information, prior beliefs and evidence collected from data. The first source of information refers to knowledge on the parameter prior to observing the data and is expressed in the form of a probability distribution. This is specifically called the prior distribution. The second source of information amounts to knowledge on observed data, which is expressed as the likelihood function of the observed data, i.e., a function of parameters given the observed data. Then, the current state of knowledge is updated by combining the two sources of information. The two sources of information are discussed below in more detail.

Prior distribution

Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_Q)$ denote a vector of Q (random) parameters whose values lie in the parameter space Θ ($\theta_q \in \Theta$). Let $p(\boldsymbol{\theta})$ denote the probability density function of the prior distribution for $\boldsymbol{\theta}$. The prior distribution is to formally incorporate any knowledge available for the parameters before observing actual data. Such knowledge originates from not only subjective beliefs but also

previous research findings. Depending on how much knowledge researchers have prior to data collection, a prior distribution is specified accordingly. Typically, there are two options for constructing a prior distribution (e.g., Carlin & Louis, 2008; Robert 2007). In some cases, we may have little information to suggest which values of the parameters might be more plausible than others. Then, we may consider using a diffuse (or uninformative) prior distribution (e.g., Box & Tiao, 1992; Press, 2003), assigning equal probabilities to a wide range of values of the parameters. A common diffuse prior is a uniform probability distribution. In the other cases, we may have considerable information about the parameters in advance, enabling the specification of so-called an informative prior distribution. When adopting a Bayesian approach to SEM, the shape of the informative prior is often chosen in such a way that the prior and posterior distributions fall into the same distributional family for a given data likelihood function (e.g., Kaplan & Depaoli, 2012; Lee & Song, 2012). Such type of the prior distribution is also known as a conjugate prior (Schlaifer & Raiffa, 1961). For instance, given that the data likelihood function is based on a normal distribution, choosing a normal prior distribution becomes a conjugate prior because it will yield a normal posterior distribution. In this dissertation, prior distributions are specified as conjugate priors, while assigning different values for the parameters of the prior distributions, specifically called hyperparameters.

Likelihood function of data

Evidence collected from the observed data is another source of information to solve for the parameters that make the occurrence of the data most likely. Such information is formally expressed by the likelihood function, defining the probability of the data being conditional on the values of $\boldsymbol{\theta}$. Suppose that g_1, \dots, g_I are independent observations of a random variable. Then, the

joint probability of $\mathbf{g} = [g_1, \dots, g_I]'$ is expressed as $p(\mathbf{g} | \boldsymbol{\theta}) = \prod_{i=1}^I p(g_i | \boldsymbol{\theta})$. This is referred to as

the likelihood function of the observed data, and the likelihood function depends on a probability distribution that is thought to give rise to the observed data. Overall, the likelihood function examines how probable we would observe the data given fixed values of the parameters.

3.2. Bayes' Theorem

As specified in the previous section, $p(\boldsymbol{\theta})$ and $p(\mathbf{g}|\boldsymbol{\theta})$ denote the probability density function of the prior distribution and the likelihood function, respectively. Let $p(\boldsymbol{\theta}|\mathbf{g})$ denote the probability density function of the posterior distribution for $\boldsymbol{\theta}$ given the observed data. $p(\boldsymbol{\theta}|\mathbf{g})$ is obtained by combining the likelihood function of the observed data with the prior distribution. In other words, it specifies how one should update the existing beliefs in the light of newly introduced evidence. The probability density function of the posterior distribution is updated via Bayes' theorem as follows:

$$p(\boldsymbol{\theta}|\mathbf{g}) = \frac{p(\mathbf{g}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\Theta} p(\mathbf{g}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (3.1)$$

Because the denominator in (3.1) is a normalizing constant that rescales $p(\mathbf{g}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ to have a proper probability distribution for $p(\boldsymbol{\theta}|\mathbf{g})$ on a $[0, 1]$, (3.1) can be rewritten as

$$p(\boldsymbol{\theta}|\mathbf{g}) \propto p(\mathbf{g}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (3.2)$$

where the symbol \propto means 'is proportional to'. From the resultant posterior distribution, one would further compute summary statistics (e.g., posterior mean and variance) by taking an expectation of a function of parameters $f(\boldsymbol{\theta})$, as follows:

$$\mathbb{E}[f(\boldsymbol{\theta})] = \int_{\Theta} f(\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{g})d\boldsymbol{\theta}, \quad (3.3)$$

where $f(\boldsymbol{\theta}) = \boldsymbol{\theta}$ for computing the mean and $f(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \mathbb{E}(\boldsymbol{\theta}))^2$ for computing the variance.

Despite the theoretical simplicity of Bayes' theorem, the practical use of a Bayesian approach was limited owing to high dimensional integrations required for obtaining a posterior distribution. That is, with two or more parameters, calculating the normalizing constant in (3.1) (or even obtaining summary statistics in (3.3)) is often intractable or cannot be carried out in closed form (e.g., Bolstad, 2007; Gilks, Richardson, & Spiegelhalter, 1996; Robert & Casella, 2004). Such intractable integration problem has remained the major obstacle in conducting Bayesian inference until the advent of Markov Chain Monte Carlo (MCMC) methods.

3.3. Markov Chain Monte Carlo (MCMC) Methods

Markov Chain Monte Carlo (MCMC) has revolutionized the application of a Bayesian approach along with the advance of computation power, making it possible to draw samples from a complex distribution of interest without actually evaluating integrations implied in the calculation. MCMC, which was developed in the early 1950s (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953), is so named because it simulates a sequence of random variables $(\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(s-1)}, \boldsymbol{\theta}^{(s)}, \boldsymbol{\theta}^{(s+1)}, \dots, \boldsymbol{\theta}^{(S)})$; where $s = 1, \dots, S$, whose values in the s th iteration $\boldsymbol{\theta}^{(s)}$ depends only on those in the previous iteration $\boldsymbol{\theta}^{(s-1)}$. This describes the 'Markov Chain' part of the term MCMC. The 'Monte Carlo' part is to denote that the simulated samples are used to approximate the integrations, as in (3.3), which are intractable in high dimensions. With the MCMC methods, adopting a Bayesian approach has become feasible even with high-dimensional integrations, because MCMC enables to not only update the posterior distribution but also construct estimators for the unknown parameters on the basis of the updated posterior distribution.

In practice, two MCMC methods are commonly used for Bayesian inference: the Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970) and Gibbs Sampler (Geman &

Geman, 1984) algorithms. The two algorithms are prevalent because they construct a Markov chain in such a way that the chain eventually reaches a unique stationary (or invariant) distribution, fulfilling the conditions of *detailed balance* and *ergodicity* (e.g., Chib & Greenberg, 1995; Gilks et al, 1996, Chapter 4; Robert & Casella, 2004, Chapter 6). *Detailed balance* is satisfied when, for every pair of two possible values of the chain, it allows the transition between the two values to be reversible with an equal probability. Namely, this ensures the Markov chain to avoid getting stuck in only one part of the distribution. *Ergodicity* is a property to ensure the existence of a unique stationary distribution independent of a starting point of the chain. In particular, a Markov chain is said to be *ergodic* when the probability of moving from one value to another is expressed always positive and the chain does not repeat an identical cycle between a set of values. Since a sequence of samples obtained from one of these MCMC methods will eventually converge to a unique stationary distribution, which is the posterior distribution in a Bayesian approach, the expectation of a function $f(\boldsymbol{\theta})$ in (3.3) would be estimated by approximating the integration with Monte Carlo integration, as follows:

$$\hat{f}(\boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^S f(\boldsymbol{\theta}^{(s)}), \quad (3.4)$$

where $\boldsymbol{\theta}^{(s)}$ is the s th iteration's sample from $p(\boldsymbol{\theta} | \mathbf{g})$, and $\hat{f}(\boldsymbol{\theta})$ is called a Monte Carlo estimator for $\boldsymbol{\theta}$.

Let $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^{(s-1)}$ denote the candidate and current values of the parameters in the Markov chain, respectively. Given the objective of constructing a Markov chain of $\boldsymbol{\theta}^{(s)}$ with $p(\boldsymbol{\theta} | \mathbf{g})$ as the desired target distribution, at the s th iteration, the Metropolis-Hastings algorithm proposes $\boldsymbol{\theta}^*$ according to a proposal distribution $q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)})$, a distribution that enables to draw new

candidate values conditional upon the current values of the parameters. An example of a proposal distribution would be a normal distribution whose mean is set to $\boldsymbol{\theta}^{(s-1)}$.

Then, the transition of the chain (i.e., a transition from $\boldsymbol{\theta}^*$ to $\boldsymbol{\theta}^{(s)}$) will occur based on a so-called acceptance probability $\rho(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)})$

$$\begin{aligned} \rho(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)}) &= \min \left\{ 1, \frac{p(\boldsymbol{\theta}^* | \mathbf{g})q(\boldsymbol{\theta}^{(s-1)} | \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(s-1)} | \mathbf{g})q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)})} \right\} \\ &= \min \{1, R\}, \end{aligned} \tag{3.5}$$

where R is called the acceptance ratio that assess the plausibility of the candidate value $\boldsymbol{\theta}^*$ relative to the current value $\boldsymbol{\theta}^{(s-1)}$. If the acceptance ratio is equal to or greater than one, the candidate value will be always accepted as $\boldsymbol{\theta}^{(s)}$, making a transition from $\boldsymbol{\theta}^*$ to $\boldsymbol{\theta}^{(s)}$. If it is smaller than one, the candidate value will be accepted based on the probability of $\rho(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)})$.

For this, a random number is drawn from a uniform distribution on the interval $[0, 1]$, and if this random value is smaller than $\rho(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)})$, then $\boldsymbol{\theta}^{(s)} = \boldsymbol{\theta}^*$. Otherwise, the Metropolis-Hastings algorithm would not move and remain the same as the current value. This is reiterated as follows:

$$\boldsymbol{\theta}^{(s)} = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \min \{1, R\} \\ \boldsymbol{\theta}^{(s-1)} & \text{otherwise.} \end{cases} \tag{3.6}$$

This process repeats until s reaches S , and the collected samples are called posterior samples.

Table 3- 2 describes a generic Metropolis-Hastings algorithm in detail.

Table 3- 2. A description of the Metropolis-Hastings algorithm.

Step 1. Generate initial value $\boldsymbol{\theta}^{(0)}$ at random, subject to $p(\boldsymbol{\theta}^{(0)} | \mathbf{g}) > 0$.

Step 2. At the s th iteration, generate a candidate value $\boldsymbol{\theta}^*$ from a proposal density function $q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)})$.

Step 3. Given the candidate value, compute the ratio of the density at the candidate ($\boldsymbol{\theta}^*$) and current ($\boldsymbol{\theta}^{(s-1)}$) values

$$R = \frac{p(\boldsymbol{\theta}^* | \mathbf{g})q(\boldsymbol{\theta}^{(s-1)} | \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(s-1)} | \mathbf{g})q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)})}$$

Step 4. Accept the candidate value according to

$$\boldsymbol{\theta}^{(s)} = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \min\{1, R\} \\ \boldsymbol{\theta}^{(s-1)} & \text{otherwise} \end{cases}$$

Step 5. Set $s = s + 1$, and repeat Steps 2 to 4 until $s = S$.

Importantly, computing the ratio of two probability density functions $\frac{p(\boldsymbol{\theta}^* | \mathbf{g})}{p(\boldsymbol{\theta}^{(s-1)} | \mathbf{g})}$ in (3.5)

cancels out the normalizing constant. Thus, the Metropolis-Hastings algorithm does not require computing the normalizing constant in (3.1), although the desired target distribution is the full posterior distribution. Also, the acceptance probability $\rho(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)})$ in (3.5) is constructed to ensure that the transition or Markov kernel, the conditional probability distribution of $\boldsymbol{\theta}^{(s-1)}$, would satisfy the *detailed balance* condition when generating a Markov chain (Chib & Greenberg, 1995; Robert & Casella, 2004, Chapter 6). Based on Chib & Greenberg (1995), the transition kernel K is expressed as

$$K(\boldsymbol{\theta}^{(s-1)}, \boldsymbol{\theta}^*) = q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)})\rho(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)}) + \tau_{\boldsymbol{\theta}^{(s-1)}}(\boldsymbol{\theta}^*)(1 - r(\boldsymbol{\theta}^{(s-1)})), \quad (3.7)$$

where $\tau_{\boldsymbol{\theta}^{(s-1)}}$ is the standard dirac-delta function (i.e., $\tau_{\boldsymbol{\theta}^{(s-1)}}$ goes to infinite when $\boldsymbol{\theta}^{(s-1)} = \boldsymbol{\theta}^*$, zero otherwise) and $r(\boldsymbol{\theta}^{(s-1)}) = \int \rho(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)})q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)})d\boldsymbol{\theta}^*$. Mathematically, the *detailed balance* is satisfied when

$$K(\boldsymbol{\theta}^{(s-1)}, \boldsymbol{\theta}^*) p(\boldsymbol{\theta}^{(s-1)} | \mathbf{g}) = K(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)}) p(\boldsymbol{\theta}^* | \mathbf{g}). \quad (3.8)$$

By multiplying $p(\boldsymbol{\theta}^{(s-1)} | \mathbf{y})$ to the first term of (3.7) and using distributive rule, (3.8) is satisfied, as follows:

$$\begin{aligned} & q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)}) \rho(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)}) p(\boldsymbol{\theta}^{(s-1)} | \mathbf{g}) \\ &= q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)}) \min \left\{ 1, \frac{p(\boldsymbol{\theta}^* | \mathbf{g}) q(\boldsymbol{\theta}^{(s-1)} | \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(s-1)} | \mathbf{g}) q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)})} \right\} p(\boldsymbol{\theta}^{(s-1)} | \mathbf{g}) \\ &= q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)}) \min \left\{ p(\boldsymbol{\theta}^{(s-1)} | \mathbf{g}), \frac{p(\boldsymbol{\theta}^* | \mathbf{y}) q(\boldsymbol{\theta}^{(s-1)} | \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)})} \right\} \\ &= q(\boldsymbol{\theta}^{(s-1)} | \boldsymbol{\theta}^*) \min \left\{ \frac{p(\boldsymbol{\theta}^{(s-1)} | \mathbf{g}) q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(s-1)})}{p(\boldsymbol{\theta}^* | \mathbf{g}) q(\boldsymbol{\theta}^{(s-1)} | \boldsymbol{\theta}^*)}, 1 \right\} p(\boldsymbol{\theta}^* | \mathbf{g}) \\ &= q(\boldsymbol{\theta}^{(s-1)} | \boldsymbol{\theta}^*) \rho(\boldsymbol{\theta}^{(s-1)}, \boldsymbol{\theta}^*) p(\boldsymbol{\theta}^* | \mathbf{g}). \end{aligned} \quad (3.9)$$

Another popular MCMC method is the Gibbs Sampler algorithm (Geman & Geman, 1984), which will be primarily adopted in this dissertation. The Gibbs Sampler algorithm is one of the most common algorithms used in a Bayesian approach to SEMs because it naturally provides a (simpler) distribution from which samples are drawn, without the need to concern about how to design a proposal distribution from the joint posterior distribution (e.g., Lee & Song, 2012; Palomo et al., 2007). The main idea behind the Gibbs Sampler algorithm is that it simplifies the simulation procedure by sampling from so-called full conditional distributions, each of which is a conditional distribution of a parameter while fixing the remaining parameters to constants. In particular, the Gibbs Sampler algorithm is a special case of the Metropolis-Hastings algorithm wherein the proposal distribution is replaced by a sequence of full conditional distributions and the acceptance ratio in (3.5) is computed to be always one (Gelman et al, 2004, Chapter 11; Robert & Casella, 2004, Chapter 7). With Q random variables

$\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_Q]$, a generic description of the Gibbs Sampler algorithm is provided in Table 3-3.

Table 3- 3. A description of the Gibbs Sampler algorithm.

Step 1. Generate initial values of parameters $\boldsymbol{\theta}^{(0)}$ at random, subject to $p(\boldsymbol{\theta}^{(0)} | \mathbf{g}) > 0$.

Step 2. At the s th iteration, successively generate new values of $\boldsymbol{\theta} = [\theta_1^{(s)}, \theta_2^{(s)}, \dots, \theta_Q^{(s)}]$ from the full conditional distributions, as follows.

Step 2-1. Draw a new value of θ_1 from $\theta_1^{(s)} \sim p(\theta_1 | \theta_2^{(s-1)}, \theta_3^{(s-1)}, \dots, \theta_Q^{(s-1)})$

Step 2-2. Continue from $q = 2, \dots, Q$ using $\theta_q^{(s)} \sim p(\theta_q | \theta_1^{(s)}, \theta_{q-1}^{(s)}, \theta_{q+1}^{(s-1)}, \dots, \theta_Q^{(s-1)})$

Step 3. Set $s = s + 1$, and repeat Step 2 until $s = S$.

As the derivation of full conditional distributions becomes more straightforward owing to the adoption of conjugate priors for parameters, the Gibbs Sampler algorithm is more frequently used in practice because it would provide posterior distributions in closed form and eliminate a step of selecting proposal distributions required in the Metropolis-Hastings algorithm. Moreover, when the number of parameters to estimate is large, as in structural equation models, the Gibbs Sampler algorithm is computationally more efficient, dividing the parameters into several subsets of them (e.g., Arminger & Muthén, 1998; Scheines et al., 1999). For these reasons, I will focus on the implementation of the Gibbs Sampler in the dissertation.

To illustrate the Gibbs sampler algorithm, let us consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \tag{3.10}$$

where y_i and x_i are the response and exploratory variables' scores of the i th observation ($i = 1, \dots, D$), respectively, β_0 is the intercept, and β_1 is the slope or regression coefficient. Assume that

$e_i \sim \text{Normal}(0, \sigma)$, where σ is the error variance. Given e_1, \dots, e_I are independent and identically distributed as $\text{Normal}(0, \sigma)$, the distribution of y_i for each \mathbf{x}_i is expressed as

$$\begin{aligned} y_i &\sim \text{Normal}(\beta_0 + \beta_1 x_i, \sigma) \\ &= \text{Normal}(\mathbf{x}_i \boldsymbol{\beta}, \sigma), \end{aligned} \quad (3.11)$$

where \mathbf{x}_i represents a matrix with one in the first column and \mathbf{x}_i 's in the second columns,

$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$. Then, the likelihood function can be written as, up to a normalizing constant,

$$p(\mathbf{y} | \boldsymbol{\beta}, \sigma) \propto \sigma^{-I/2} \exp \left\{ -\frac{1}{2\sigma} \sum_{i=1}^I (y_i - \mathbf{x}_i \boldsymbol{\beta})^2 \right\}. \quad (3.12)$$

To discuss a Bayesian approach for fitting the simple linear regression model, prior distributions for the parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma\}$ must be specified. In this example, we specify the distribution for $\boldsymbol{\beta}$ as being independent of σ (i.e., $p(\boldsymbol{\beta}, \sigma) = p(\boldsymbol{\beta}) p(\sigma)$), and consider the conjugate prior distributions for both sets of parameters, so as to have full conditional distributions in the same distributional family as the posterior distributions. The prior distributions are specified as follows:

$$\begin{aligned} p(\boldsymbol{\beta}) &\sim \text{Normal}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \\ p(\sigma) &\sim \text{IG}(a, \nu), \end{aligned} \quad (3.13)$$

where *IG* stands for an inverse gamma distribution with the shape parameter a and the scale parameter ν . Following Bayes' theorem in (3.1), the joint posterior distribution is expressed as

$$p(\boldsymbol{\beta}, \sigma | \mathbf{y}) = \frac{p(\mathbf{y} | \boldsymbol{\beta}, \sigma) p(\boldsymbol{\beta}) p(\sigma)}{\int p(\mathbf{y} | \boldsymbol{\beta}, \sigma) p(\boldsymbol{\beta}) p(\sigma) d\boldsymbol{\beta} d\sigma}, \quad (3.14)$$

Under this model, combining (3.12) and (3.13), the probability density function of the joint posterior distribution is obtained, up to a normalizing constant, as

$$p(\boldsymbol{\beta}, \sigma | \mathbf{y}) \propto \sigma^{-(I/2+a+1)} \exp \left\{ -\frac{1}{2\sigma} \left[2\nu + \sum_{i=1}^I (y_i - \mathbf{x}_i \boldsymbol{\beta})^2 \right] + (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)' \boldsymbol{\Sigma}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta) \right\}. \quad (3.15)$$

Here, the Gibbs Sampler algorithm can be used because sampling from the full conditional distributions is much easier and quicker than drawing directly from a complex joint posterior distribution expressed in (3.15). As such, (3.15) is further simplified with the full conditional distributions $p(\boldsymbol{\beta} | \sigma, \mathbf{y})$ and $p(\sigma | \boldsymbol{\beta}, \mathbf{y})$. The full conditional distribution for $\boldsymbol{\beta}$ is obtained by rearranging (3.15) as follows:

$$p(\boldsymbol{\beta} | \sigma, \mathbf{y}) \sim \text{Normal} \left(\boldsymbol{\Omega}_\beta^{-1} \left(\frac{1}{\sigma} \sum_{i=1}^I \mathbf{x}_i' y_i + \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\mu}_\beta \right), \boldsymbol{\Omega}_\beta^{-1} \right), \quad (3.16)$$

where $\boldsymbol{\Omega}_\beta = \left(\frac{1}{\sigma} \sum_{i=1}^I \mathbf{x}_i' \mathbf{x}_i + \boldsymbol{\Sigma}_\beta^{-1} \right)$. Likewise, the full conditional distribution for σ is given by

$$p(\sigma | \boldsymbol{\beta}, \mathbf{y}) \sim \text{IG} \left(a + \frac{I}{2}, \nu + \frac{1}{2} \sum_{i=1}^I (y_i - \mathbf{x}_i \boldsymbol{\beta})^2 \right). \quad (3.17)$$

Thus, the Gibbs Sampler algorithm can proceed as follows: at the s th iteration, it draws $\boldsymbol{\beta}^{(s)}$ from $p(\boldsymbol{\beta} | \sigma^{(s-1)}, \mathbf{y})$ and subsequently draws $\sigma^{(s)}$ from $p(\sigma | \boldsymbol{\beta}^{(s)}, \mathbf{y})$.

Although the abovementioned algorithm ensures convergence to the target distribution in theory, its successful implementation depends on the number of iterations (S). This is because we initialize the starting values with random values, so that the posterior samples may not necessarily guarantee the convergence especially when the number of iterations is not large enough. To examine if a sequence of posterior samples obtained from an MCMC method converges to the target distribution after S iterations, we would draw trace plots for the samples against S iterations. When the values of the samples move up and down around the mean of the distribution but with a relative small amount of fluctuations, we may conclude that the samples have reached the target distribution sufficiently well. On the other hand, when a lack of

convergence is suspected by showing a systematic trend over the iterations (e.g., staying in a certain range of values for a longer period of time rather than traversing up and down), we would need to increase the total number of S iterations. Alternatively, we would increase the number of early iterations to discard in the chain (i.e., burn-in sample size). In practice, first 1000 to 5000 samples are thrown out. Moreover, an autocorrelation between the generated samples can be used as another diagnostic measure for assessing convergence (e.g., Lynch, 2007). The autocorrelation among the obtained MCMC samples for the q th parameter of interest (θ_q) over lag U is computed as the correlation between the value of every sample and that of the U th lag as follows:

$$\text{ACF}_U = \frac{\sum_{s=1}^{S-U} (\theta_q^{(s)} - \bar{\theta}_q)(\theta_q^{(s+U)} - \bar{\theta}_q)}{\sum_{s=1}^S (\theta_q^{(s)} - \bar{\theta}_q)^2}, \quad (3.18)$$

where $\bar{\theta}_q$ represent the mean of all sampled values for θ_q . We expect the autocorrelation over lag U to be smaller as the value of U increases. When a high autocorrelation is present, we need to increase the value of U , also called thinning, beyond which the autocorrelation becomes small enough to ignore, preferably very close to zero.

In addition to obtaining a posterior distribution for parameters, a Bayesian approach enables to have a predictive distribution for making inferences about an unknown but potentially observable quantity, such as future observations (Congdon, 2007; Gelman et al., 2004, Chapter 6). Denoting \mathbf{g}^{rep} as the replicated future observations, the probability density function for the replicated data is called the posterior predictive distribution, and is expressed as

$$\begin{aligned} p(\mathbf{g}^{rep} | \mathbf{g}) &= \int p(\mathbf{g}^{rep} | \boldsymbol{\theta}) p(\mathbf{g} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int p(\mathbf{g}^{rep} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{g}) d\boldsymbol{\theta}. \end{aligned} \quad (3.19)$$

When an MCMC method is employed to generate $\boldsymbol{\theta}$ from the posterior distribution, we may simulate \mathbf{g}^{rep} from the posterior predictive distribution as a part of the MCMC method. Given values of $\boldsymbol{\theta}^{(s)}$ obtained via an MCMC method at the s th iteration, a new observation \mathbf{g}^{rep} is generated from the likelihood function of the observed data $p(\mathbf{g} | \boldsymbol{\theta}^{(s)})$. This posterior predictive distribution would be of particular use for evaluating the overall fit of a hypothesized model, which will be discussed in the following section.

3.4. Model Assessment and Comparison in Bayesian Inference

It is important to assess the plausibility of a hypothesized model, examining if the model provides a reasonable summary of the data. For this, so-called posterior predictive checking proposed by Gelman, Meng, and Stern (1996) can be adopted. The basic idea of the posterior predictive checking is to replicate data under the hypothesized model and compare them with the original observed data. If the model fits well to the observed data, there would be little discrepancy between the replicated and observed data. To formally quantify this model fit, a set of replicated data \mathbf{g}^{rep} is first generated based on $\boldsymbol{\theta}^{(s)}$, being of the same size as the original data, under the fitted model M . Then, the chi-square statistic based on the likelihood function is computed to measure discrepancy for both \mathbf{g}^{rep} and \mathbf{g} , denoted by $D(\mathbf{g}^{rep}; \boldsymbol{\theta}^{(s)})$ and $D(\mathbf{g}; \boldsymbol{\theta}^{(s)})$, respectively. Having obtained $\left\{ \left(D(\mathbf{g}^{rep}; \boldsymbol{\theta}^{(s)}), D(\mathbf{g}; \boldsymbol{\theta}^{(s)}) \right), s=1, \dots, S \right\}$, we compute a Bayesian posterior predictive p -value (PPP_b) (e.g., Lynch, 2007; Scheines et al., 1999) as follows:

$$PPP_b = p \left[D(\mathbf{g}^{rep}; \boldsymbol{\theta}) \geq D(\mathbf{g}; \boldsymbol{\theta}) \mid M, \mathbf{g} \right] \approx \frac{1}{S} \sum_{s=1}^S \varpi_s, \quad (3.20)$$

where $\varpi_s = 1$ if $D(\mathbf{g}^{rep}; \boldsymbol{\theta}^{(s)}) \geq D(\mathbf{g}; \boldsymbol{\theta}^{(s)})$ and 0 otherwise. Overall, PPP_b provides a proportion of the discrepancy measure of the replicated data equal to or greater than that of the actual observed data. A PPP_b value around 0.50 indicates a good fit, whereas an extreme value (e.g., value close to either 0 or 1) suggests a bad fit.

Besides evaluating the adequacy of a single model, researchers may also be interested in comparing two or more models that are all theoretically plausible. In a Bayesian approach, such model comparison is addressed using the Bayes factor (Berger, 1985; Kass & Raftery, 1995) or its variants. The Bayes factor is considered the Bayesian counterpart of classical hypothesis testing that is based on p -values determined by asymptotic distributions of a test statistic. In this section, we introduce the Bayes factor for model comparison, assessing which model accounts for data better among two or more competing models. If there are two competing models for observed data \mathbf{g} , say M_0 and M_1 , the Bayes factor provides a quantity that expresses the extent to which \mathbf{g} supports M_0 over M_1 by taking the ratio of two integrated likelihoods (Kass & Raftery, 1995). Specifically, the Bayes factor for M_0 against M_1 , denoted as K_{01} , is given as

$$K_{01} = \frac{p(\mathbf{g} | M_0)}{p(\mathbf{g} | M_1)}, \quad (3.21)$$

where the marginal likelihood $p(\mathbf{g} | M_k)$ indicates the probability of observing \mathbf{g} given the k th model, integrating over $\boldsymbol{\theta}_k$, given by $p(\mathbf{g} | M_0) = \int p(\mathbf{g} | \boldsymbol{\theta}_0, M_0) p(\boldsymbol{\theta}_0 | M_0) d\boldsymbol{\theta}_0$ and $p(\mathbf{g} | M_1) = \int p(\mathbf{g} | \boldsymbol{\theta}_1, M_1) p(\boldsymbol{\theta}_1 | M_1) d\boldsymbol{\theta}_1$, in which $p(\mathbf{g} | \boldsymbol{\theta}_k, M_k)$ indicates a probability density function of data parameterized by $\boldsymbol{\theta}_k$ in the k th model. Note that this marginal likelihood is equivalent to the normalizing constant used in the Bayes' theorem in (3.2).

Nevertheless, as stated earlier, $p(\mathbf{g} | M_k)$ is often hard to evaluate analytically when it requires integration over a large number of parameters. As a result, several alternative methods had been proposed to approximate the marginal likelihood: Monte Carlo estimator via importance sampling (Newton & Raftery, 1994; Gelfand & Dey, 1994), Laplace-Metropolis estimator (Lewis & Raftery, 1997), Chib's estimator (Chib, 1995), bridge sampling estimator (Meng & Wong, 1996), path sampling estimator (Gelman & Meng, 1998), and annealed importance sampling estimator (Neal, 2001).

Of these methods, in this dissertation, I focus on Chib's estimator because it is easily computable based on the results obtained from the Gibbs Sampler algorithm and flexible for accommodating a various range of models, from the simplest case with a single parameter to more general cases with more than two parameters. To illustrate how this method works, let us assume that there are two unknown parameters, i.e., $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Chib (1995) had introduced an efficient method that approximates the marginal likelihood by substituting the posterior means computed from the Gibbs output, denoted by $\boldsymbol{\theta}^\# = (\theta_1^\#, \theta_2^\#)$, for the parameters and rearranging the Bayes' theorem in (3.1), on the logarithm scale, as follows:

$$p(\mathbf{y}) = \ln p(\mathbf{g} | \boldsymbol{\theta}^\#) + \ln p(\boldsymbol{\theta}^\#) - \ln p(\boldsymbol{\theta}^\# | \mathbf{g}). \quad (3.22)$$

Having two unknown parameters, the third term in (3.22) is rewritten as

$$\begin{aligned} p(\boldsymbol{\theta}^\# | \mathbf{g}) &= p(\theta_1^\#, \theta_2^\# | \mathbf{g}) \\ &= p(\theta_1^\# | \mathbf{g}) p(\theta_2^\# | \mathbf{g}, \theta_1^\#). \end{aligned} \quad (3.23)$$

Then, by obtaining the posterior draws of θ_2 (i.e., $\{\theta_2^{(1)}, \theta_2^{(2)}, \dots, \theta_2^{(S)}\}$ from the Gibbs Sampler algorithm over S iterations), a Monte Carlo estimator of the first term in the right side of (3.23) is updated by

$$\hat{p}(\theta_1^\# | \mathbf{g}) = \frac{1}{S} \sum_{s=1}^S p(\theta_1^\# | \mathbf{g}, \theta_2^{(s)}). \quad (3.24)$$

Similarly, the second term in the right side of (3.23) is estimated by implementing another Gibbs Sampler algorithm to draw samples of θ_2 , but now being conditional on $\theta_1^\#$ rather than the s th iteration's output of θ_1 . Then, $p(\theta_2^\# | \mathbf{g}, \theta_1^\#)$ is estimated by

$$\hat{p}(\theta_2^\# | \mathbf{g}, \theta_1^\#) = \frac{1}{S} \sum_{s=1}^S p(\theta_2^\# | \mathbf{g}, \theta_1^\#). \quad (3.25)$$

The Monte Carlo estimator obtained from (3.24) and (3.25) are used to compute (3.23) and consequently the Bayes factor in (3.21).

Bayesian information criterion (BIC), also called Schwarz criterion (Schwarz, 1978), is a measure derived by approximating the Bayes factor. This approximation is appealing in many Bayesian modeling problems especially when the integrated likelihoods are computationally hard to evaluate or there is lack of appropriate priors for the parameters (Kass & Raftery, 1995; Kass & Wasserman, 1995). The BIC is defined as $\text{BIC} = -2 \ln \mathbf{L}^* + q \log(I)$, where $\ln \mathbf{L}^*$ is the maximized value of a log-likelihood function of data given estimated values for parameters and a model (i.e., $\mathbf{L}^* = p(\mathbf{g} | \boldsymbol{\theta}, M)$), and q is the number of parameters to be estimated. The second term on the right-hand side of the BIC is to handle an over-fitting problem by imposing a penalty on the number of parameters. When the objective of study is to compare several models, the difference between two BICs can be computed as

$$\Delta \text{BIC}_{01} = \text{BIC}_{M_0} - \text{BIC}_{M_1} = \ln \mathbf{L}_{M_0}^* - \ln \mathbf{L}_{M_1}^* - \frac{1}{2}(q_0 - q_1) \log(I), \quad (3.26)$$

where q_0 and q_1 indicate the numbers of free parameters in M_0 and M_1 , respectively. If the BIC difference is greater than six, there is strong evidence that the model with the smaller BIC would be favored (Kass & Raftery, 1995).

Deviance information criterion (DIC) (Spiegelhalter et al., 2002) is another information criterion used for model comparison. The DIC for a single model is computed as

$$\text{DIC} = \mathbf{p}_\Delta + \bar{\Delta}, \quad (3.27)$$

where $\mathbf{p}_\Delta = \bar{\Delta} - \Delta(\bar{\boldsymbol{\theta}})$ and $\bar{\Delta} = \mathbb{E}_{\boldsymbol{\theta}} \{\Delta(\boldsymbol{\theta})\}$, in which $\Delta(\boldsymbol{\theta}) = -2\log(\mathbf{g} | \boldsymbol{\theta}) + 2\log(h(\mathbf{g}))$ is a deviance measure, $\bar{\boldsymbol{\theta}}$ is the posterior mean of the parameters, $\mathbb{E}_{\boldsymbol{\theta}} \{\cdot\}$ refers to a conditional expectation of the deviance statistic given the parameters, and $h(\mathbf{g})$ is a standardizing factor irrelevant to the parameters in a given model, which will be cancelled out when comparing models. A model with a smaller DIC value is preferred to other models with larger DIC values.

Chapter 4. Bayesian Generalized Structured Component Analysis

The previous two chapters discussed the technical underpinnings of the original GSCA and Bayesian inference. In this chapter, Bayesian GSCA or simply BGSCA is developed by integrating the original GSCA into a Bayesian framework. The chapter begins with model specification of BGSCA, followed by its parameter estimation procedure. It also discusses other possible extensions of BGSCA.

4.1. Model Specification

Let \mathbf{z}_i denote a (J_1+J_2) by 1 vector of observed variables, as defined in (2.2). Let $\bar{\mathbf{z}}_i = \begin{bmatrix} \bar{\mathbf{x}}_i \\ \bar{\mathbf{y}}_i \end{bmatrix}$ denote a (J_1+J_2) by 1 vector of so-called error-free variables, where $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{y}}_i$ are J_1 by 1 and J_2 by 1 vector of the error-free counterparts of exogenous and endogenous observed variables, \mathbf{x}_i and \mathbf{y}_i , respectively. The error-free variables $\bar{\mathbf{z}}_i$ contain true underlying scores or values of \mathbf{z}_i (Nounou, Bakshi, Goel, & Shen, 2002; Wentzell, Andrews, Hamilton, Faber, & Kowalski, 1997), and the observed variables \mathbf{z}_i are thus assumed to be the realizations of the true measurements perturbed by additive noises or random errors. Hereafter, the double bar symbol ($\bar{=}$) is used to indicate the true or error-free counterparts of variables or parameters in original GSCA.

Mathematically, this relationship can be expressed as

$$\mathbf{z}_i = \bar{\mathbf{z}}_i + \boldsymbol{\varepsilon}_i \quad (4.1)$$

where $\boldsymbol{\varepsilon}_i = \begin{bmatrix} \boldsymbol{\varepsilon}_i^{(x)} \\ \boldsymbol{\varepsilon}_i^{(y)} \end{bmatrix}$ is a (J_1+J_2) by 1 of the noise/error terms. Note that (4.1) was also

contemplated in maximum likelihood principal component analysis (MLPCA; Tipping & Bishop, 1999; Wentzell et al., 1997) and Bayesian principal component analysis (BPCA; Bishop, 1999; Nounou et al., 2002).

As in BPCA, BGSCA defines latent variables as weighted composites of the error-free variables $\bar{\bar{\mathbf{z}}}_i$ rather than the observed variables \mathbf{z}_i . Because BGSCA enables to take into account information about the noise/error terms when constructing latent variables, this would be of particular use when the sizes of the noise terms are perceived to be relatively large. Accounting for the noise terms in all observed variables, BGSCA modifies the weighted relation model in (2.4) as follows:

$$\bar{\bar{\mathbf{\Gamma}}}_i = \bar{\bar{\mathbf{W}}}^{\prime} \bar{\bar{\mathbf{z}}}_i, \quad (4.2)$$

As $\bar{\bar{\mathbf{z}}}_i$ is of the same size as \mathbf{z}_i , $\bar{\bar{\mathbf{\Gamma}}}_i = \begin{bmatrix} \bar{\bar{\gamma}}_i^{(x)} \\ \bar{\bar{\gamma}}_i^{(y)} \end{bmatrix}$ is a (P_1+P_2) by 1 vector of the latent variable scores

that are defined as the composites of $\bar{\bar{\mathbf{z}}}_i$, where $\bar{\bar{\gamma}}_i^{(x)}$ and $\bar{\bar{\gamma}}_i^{(y)}$ are a P_1 by 1 and P_2 by 1 vector of exogenous and endogenous error-free latent variable scores for the i th subject, respectively. Let

$\bar{\bar{\mathbf{W}}} = \begin{bmatrix} \bar{\bar{\mathbf{W}}}^{(x)} & \mathbf{0} \\ \mathbf{0} & \bar{\bar{\mathbf{W}}}^{(y)} \end{bmatrix}$ denote a (J_1+J_2) by (P_1+P_2) matrix of weights assigned to $\bar{\bar{\mathbf{z}}}_i$. From (4.2), the

measurement model of BGSCA is specified as

$$\mathbf{z}_i = \mathbf{C}^{\prime} \bar{\bar{\mathbf{\Gamma}}}_i + \boldsymbol{\delta}_i, \quad (4.3)$$

where $\mathbf{C}^{\prime} = \begin{bmatrix} \mathbf{C}^{(x)\prime} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{(y)\prime} \end{bmatrix}$ and $\boldsymbol{\delta}_i = \begin{bmatrix} \boldsymbol{\delta}_i^{(x)} \\ \boldsymbol{\delta}_i^{(y)} \end{bmatrix}$ contains the measurement errors for the observed

variables, taking into account the unexplained variance of the observed variables by the error-free latent variables.

Let \mathbf{H} denote a P_2 by P_1 matrix relating endogenous to exogenous latent variables and \mathbf{N} denote a P_2 by P_2 matrix relating endogenous latent variables to other endogenous variables.

Combining those two matrices, $\mathbf{V} = [\mathbf{H} \quad \mathbf{N}]$ is a P_2 by $(P_1 + P_2)$ matrix of path coefficients of all latent variables affecting $\gamma_i^{(y)}$. Let ζ_i denote a P_2 by 1 vector of errors of endogenous latent variables. Then, the structural model of BGSCA is expressed as

$$\gamma_i^{(y)} = \mathbf{V}\bar{\Gamma}_i + \zeta_i. \tag{4.4}$$

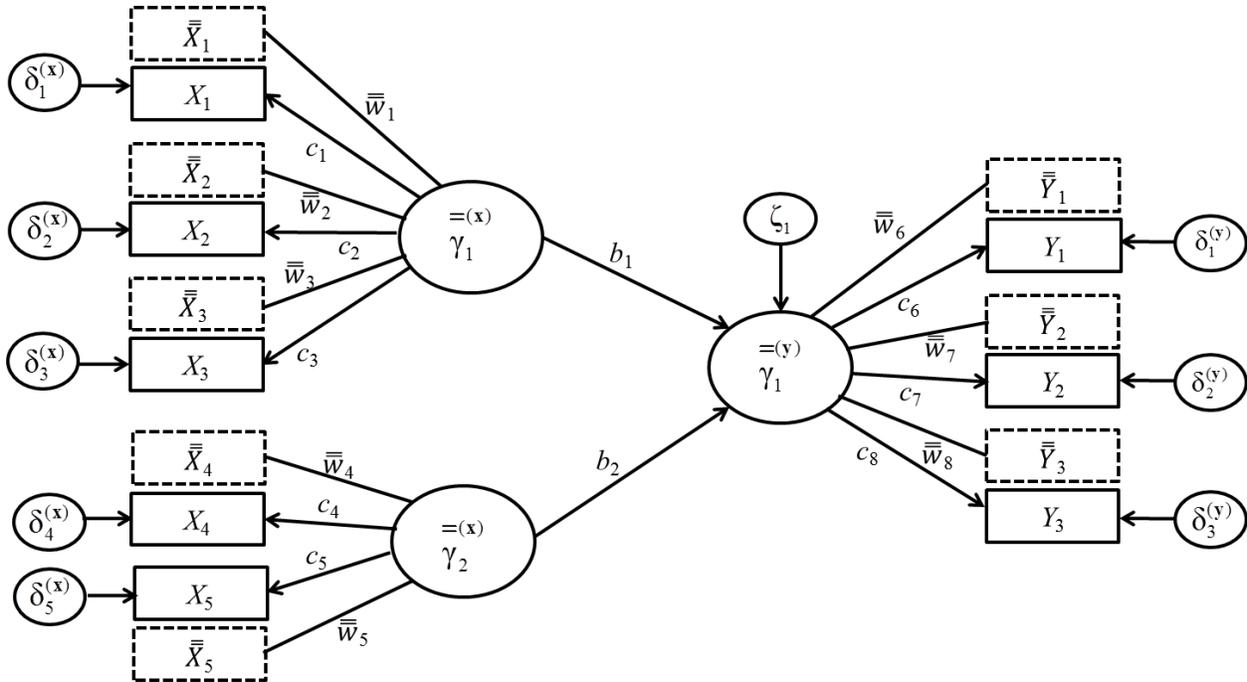


Figure 4- 1. A hypothetical BGSCA model.

To illustrate model specification of BGSCA, let us use the same hypothetical example introduced in Section 2.1. Figure 4- 1 displays the hypothetical model with the error-free counterparts of the observed variables, which are represented as rectangles in dashed line. The weighted relation model for this example is expressed as follows:

$$\begin{bmatrix} \gamma_{i1}^{(x)} \\ \gamma_{i2}^{(x)} \\ \gamma_{i1}^{(y)} \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & w_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & w_4 & w_5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_6 & w_7 & w_8 \end{bmatrix} \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \\ x_{i5} \\ y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix}, \quad (4.5)$$

$$\bar{\Gamma}_i = \bar{\mathbf{W}}' \bar{\mathbf{z}}_i.$$

The measurement model is specified as

$$\begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \\ x_{i5} \\ y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix} = \begin{bmatrix} c_1 & 0 & 0 \\ c_2 & 0 & 0 \\ c_3 & 0 & 0 \\ 0 & c_4 & 0 \\ 0 & c_5 & 0 \\ 0 & 0 & c_6 \\ 0 & 0 & c_7 \\ 0 & 0 & c_8 \end{bmatrix} \begin{bmatrix} \gamma_{i1}^{(x)} \\ \gamma_{i2}^{(x)} \\ \gamma_{i1}^{(y)} \end{bmatrix} + \begin{bmatrix} \delta_{i1}^{(x)} \\ \delta_{i2}^{(x)} \\ \delta_{i3}^{(x)} \\ \delta_{i4}^{(x)} \\ \delta_{i1}^{(y)} \\ \delta_{i2}^{(y)} \\ \delta_{i3}^{(y)} \end{bmatrix}, \quad (4.6)$$

$$\mathbf{z}_i = \mathbf{C}' \bar{\Gamma}_i + \boldsymbol{\delta}_i,$$

In Figure 4- 1, it is assumed that the error-free endogenous latent variable $\gamma_{i1}^{(y)}$ is affected by two error-free exogenous latent variables $\gamma_{i1}^{(x)}$ and $\gamma_{i2}^{(x)}$. Thus, the structural model is expressed as

$$\gamma_{i1}^{(y)} = [b_1 \quad b_2 \quad 0] \begin{bmatrix} \gamma_{i1}^{(x)} \\ \gamma_{i2}^{(x)} \\ \gamma_{i1}^{(y)} \end{bmatrix} + \zeta_{i1}, \quad (4.7)$$

$$\bar{\gamma}_i^{(y)} = \mathbf{V} \bar{\Gamma}_i + \zeta_i.$$

4.2. Parameter Estimation

To estimate unknown parameters in the BGSCA model, the likelihood function of the observed data and prior distributions should be specified from the outset, as discussed in Section 3.1. The construction of the likelihood function is dependent on how we specify the distributions for the error terms in the measurement and structural models. I here follow the specification of distributions typically contemplated in linear factor-based SEMs (e.g., Bollen, 1989).

Specifically, in the measurement model, the error vectors $\delta_i^{(x)}$ and $\delta_i^{(y)}$ are assumed to be independent of the latent variables $\bar{\Gamma}_i$ and the structural error term ζ_i . They are also assumed to follow a normal distribution.

$$\begin{aligned}\delta_i^{(x)} &\sim \text{Normal}(\mathbf{0}, \Sigma_x) \\ \delta_i^{(y)} &\sim \text{Normal}(\mathbf{0}, \Sigma_y),\end{aligned}\tag{4.8}$$

where $\Sigma_x = \text{diag}(\phi_1^{(x)}, \dots, \phi_{J_1}^{(x)})$ is a J_1 by J_1 diagonal matrix and $\Sigma_y = \text{diag}(\phi_1^{(y)}, \dots, \phi_{J_2}^{(y)})$ is a J_2 by J_2 diagonal matrix. Assuming such diagonal matrices is common but not required. In the structural model, the P_2 by 1 vector of the errors of endogenous latent variables ζ_i is assumed to be independent of exogenous latent variables $\bar{\gamma}_i^{(x)}$ and to follow a normal distribution:

$$\zeta_i \sim \text{Normal}(\mathbf{0}, \Psi),\tag{4.9}$$

where $\Psi = \text{diag}(\phi_1^{(y)}, \dots, \phi_{P_2}^{(y)})$ is a P_2 by P_2 diagonal matrix. Given the specifications of its sub-models and distributional assumptions, overall, BGSCA involves five sets of parameters to be estimated: the scores of error-free observed variables ($\bar{\mathbf{z}}_i$), weights ($\bar{\mathbf{W}}$), loadings (\mathbf{C}), path coefficients (\mathbf{V}), and the variances of the error terms in the measurement and structural models

($\phi_1^{(x)}, \dots, \phi_{J_1}^{(x)}, \phi_1^{(y)}, \dots, \phi_{J_2}^{(y)}$, and $\phi_1^{(\gamma)}, \dots, \phi_{P_2}^{(\gamma)}$). Tables 4- 1 and 4- 2 summarize notations of all variables and parameters in BGSCA, along with distributional assumptions on all the error terms.

Table 4- 1. Summary of notations for the variables in BGSCA.

Symbol	Dimension	Definition
Error-free observed random variables		
$\bar{\mathbf{x}}_i$	$J_1 \times 1$	Error-free observed exogenous variables
$\bar{\mathbf{y}}_i$	$J_2 \times 1$	Error-free observed endogenous variables
$\bar{\mathbf{z}}_i$	$(J_1+J_2) \times 1$	All error-free observed variables,
Latent random variables		
$\bar{\bar{\mathbf{y}}}_i^{(x)}$	$P_1 \times 1$	Latent exogenous variables
$\bar{\bar{\mathbf{y}}}_i^{(y)}$	$P_2 \times 1$	Latent endogenous variables
$\bar{\bar{\mathbf{\Gamma}}}_i$	$(P_1+P_2) \times 1$	All latent variables, i.e., $\bar{\bar{\mathbf{\Gamma}}}_i = \left[\bar{\bar{\mathbf{y}}}_i^{(x)}; \bar{\bar{\mathbf{y}}}_i^{(y)} \right]$
Error random variables		
$\delta_i^{(x)}$	$J_1 \times 1$	Measurement error terms in \mathbf{x}_i
$\delta_i^{(y)}$	$J_2 \times 1$	Measurement error terms in \mathbf{y}_i
δ_i	$(J_1+J_2) \times 1$	Measurement error terms in all observed variables (\mathbf{z}_i)
ζ_i	$P_2 \times 1$	Structural error terms for $\bar{\bar{\mathbf{y}}}_i^{(y)}$
Model parameters		
$\bar{\bar{\mathbf{W}}}^{(x)}$	$J_1 \times P_1$	Component weights applied to \mathbf{x}_i
$\bar{\bar{\mathbf{W}}}^{(y)}$	$J_2 \times P_2$	Component weights applied to \mathbf{y}_i
$\mathbf{C}^{(x)}$	$P_1 \times J_1$	Loadings relating $\bar{\bar{\mathbf{y}}}_i^{(x)}$ to \mathbf{x}_i
$\mathbf{C}^{(y)}$	$P_2 \times J_2$	Loadings relating $\bar{\bar{\mathbf{y}}}_i^{(y)}$ to \mathbf{y}_i
\mathbf{V}	$P_2 \times (P_1+ P_2)$	Path coefficients in the structural model
$\Sigma_x(\phi_{j_1}^{(x)})$	$J_1 \times J_1$	Measurement error variances for \mathbf{x}_i
$\Sigma_y(\phi_{j_2}^{(y)})$	$J_2 \times J_2$	Measurement error variances for \mathbf{y}_i
$\Psi(\phi_{p_2}^{(\gamma)})$	$P_2 \times P_2$	Latent error variances

Table 4- 2. Specification of BGSCA sub-models and distributional assumptions.

Measurement model

$$\mathbf{x}_i = \mathbf{C}^{(x)'} \boldsymbol{\gamma}_i^{(x)} + \boldsymbol{\delta}_i^{(x)} \quad \boldsymbol{\delta}_i^{(x)} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_x)$$

$$\mathbf{y}_i = \mathbf{C}^{(y)'} \boldsymbol{\gamma}_i^{(y)} + \boldsymbol{\delta}_i^{(y)} \quad \boldsymbol{\delta}_i^{(y)} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_y)$$

Or

$$\mathbf{z}_i = \mathbf{C}' \bar{\boldsymbol{\Gamma}}_i + \boldsymbol{\delta}_i \quad \boldsymbol{\delta}_i \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_z), \text{ where } \boldsymbol{\Sigma}_z = \begin{bmatrix} \boldsymbol{\Sigma}_x & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_y \end{bmatrix}$$

Structural model

$$\boldsymbol{\gamma}_i^{(y)} = \mathbf{V} \bar{\boldsymbol{\Gamma}}_i + \boldsymbol{\zeta}_i. \quad \boldsymbol{\zeta}_i \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Psi})$$

Weighted relation model

$$\bar{\boldsymbol{\Gamma}}_i = \bar{\mathbf{W}}' \mathbf{z}_i$$

Unlike the original GSCA, BGSCA is capable of deriving the likelihood function of the observed data. Let $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_I]'$ denote an I by (J_1+J_2) matrix of observed variables, where $\mathbf{z}_1, \dots, \mathbf{z}_I$ are independent and identically distributed. Then, the likelihood function is given by

$$p(\mathbf{Z} | \bar{\mathbf{Z}}, \boldsymbol{\theta}) = \prod_{i=1}^I \text{Normal}(\mathbf{z}_i | \bar{\mathbf{z}}_i, \boldsymbol{\theta}), \quad (4.10)$$

where $\boldsymbol{\theta} = \{\bar{\mathbf{W}}, \mathbf{C}, \mathbf{V}, \boldsymbol{\Sigma}_z, \boldsymbol{\Psi}\}$. Since the measurement error vectors $\boldsymbol{\delta}_i^{(x)}$ and $\boldsymbol{\delta}_i^{(y)}$ are assumed to be independent of each other, the likelihood function of the observed data in (4.10) can be rewritten as the product of the individual likelihood functions of $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_I]'$ and $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_I]'$ as

$$p(\mathbf{Z} | \bar{\mathbf{Z}}, \boldsymbol{\theta}) = p(\mathbf{X}, \mathbf{Y} | \bar{\mathbf{X}}, \bar{\mathbf{Y}}, \boldsymbol{\theta}) = \prod_{i=1}^I p(\mathbf{x}_i, \mathbf{y}_i | \bar{\mathbf{x}}_i, \bar{\mathbf{y}}_i, \boldsymbol{\theta}). \quad (4.11)$$

For parameter estimation in BGSCA, conjugate prior distributions are specified with different values of hyperparameters, depending on prior information available from analyses of similar or previous data. An advantage of using such conjugate prior distributions is that it simplifies the process of deriving full conditional distributions required in the Gibbs Sampler algorithm. Because the number of parameters to update in $\boldsymbol{\theta}$ is typically large in SEM, it is customary to employ the Gibbs Sampler algorithm and further specify conjugate prior distributions to obtain a posterior distribution in closed form (Kaplan & Depaoli, 2012; Palomo et al., 2007; Scheines et al., 1999). More specifically, given the distribution form of the likelihood function in (4.10), we choose normal priors for $\bar{\mathbf{z}}_i$, $\bar{\mathbf{W}}$, \mathbf{C} , and \mathbf{V} . Inverse gamma distributions are specified as the prior distribution for $\phi_1^{(x)}, \dots, \phi_{J_1}^{(x)}, \phi_1^{(y)}, \dots, \phi_{J_2}^{(y)}$, and $\phi_1^{(\gamma)}, \dots, \phi_{p_2}^{(\gamma)}$, which are consistent with what has been assumed in Bayesian linear regression models (e.g., Carlin & Louis, 2008) or Bayesian factor-based SEM (e.g., Muthén & Asparouhov, 2012). The specification of the conjugate prior distributions will be delineated below.

The Gibbs Sampler algorithm for BGSCA consists of several steps, at each of which, a set of parameters is sampled from the corresponding full conditionals.

Step 1. This step generates the error-free variables $\bar{\mathbf{z}}_i$, given the observed data \mathbf{z}_i and the parameters. As $\delta_i^{(x)}$ and $\delta_i^{(y)}$ are assumed to be independent of each other, the error-free variables can be estimated by using the individual likelihood functions of \mathbf{x}_i and \mathbf{y}_i . The prior distribution for $\bar{\mathbf{x}}_i$ is specified to follow a normal distribution with a J_1 by 1 mean vector \mathbf{m}_x and J_1 by (J_1+J_2) covariance matrix \mathbf{Q}_x , i.e., $\bar{\mathbf{x}}_i \sim Normal(\mathbf{m}_x, \mathbf{Q}_x)$. This prior is known to be a reasonable and common choice for the error-free variables (e.g., Nounou et al., 2002). Combining the prior

distribution with the likelihood function, the full conditional distribution for $\bar{\mathbf{X}} = [\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_I]'$ is

derived as:

$$\begin{aligned}
 p(\bar{\mathbf{X}} | \mathbf{X}, \boldsymbol{\theta}) &\propto \prod_{i=1}^I p(\mathbf{x}_i | \bar{\mathbf{x}}_i, \boldsymbol{\theta}) p(\boldsymbol{\gamma}_i^{(y)} | \bar{\mathbf{x}}_i, \bar{\mathbf{w}}^{(x)}, \mathbf{V}) p(\bar{\mathbf{x}}_i) \\
 &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^I \left(\mathbf{x}_i - \mathbf{C}^{(x)'} \bar{\mathbf{W}}^{(x)} \bar{\mathbf{x}}_i \right)' \boldsymbol{\Sigma}_x^{-1} \left(\mathbf{x}_i - \mathbf{C}^{(x)'} \bar{\mathbf{W}}^{(x)} \bar{\mathbf{x}}_i \right) \right. \\
 &\quad \left. + \left(\boldsymbol{\gamma}_i^{(y)} - \mathbf{N} \boldsymbol{\gamma}_i^{(y)} - \mathbf{H} \boldsymbol{\gamma}_i^{(x)} \right)' \boldsymbol{\Psi}^{-1} \left(\boldsymbol{\gamma}_i^{(y)} - \mathbf{N} \boldsymbol{\gamma}_i^{(y)} - \mathbf{H} \boldsymbol{\gamma}_i^{(x)} \right) + \left(\bar{\mathbf{x}}_i - \mathbf{m}_x \right)' \mathbf{Q}_x^{-1} \left(\bar{\mathbf{x}}_i - \mathbf{m}_x \right) \right\}.
 \end{aligned} \tag{4.12}$$

Rearranging (4.12) with respect to $\bar{\mathbf{x}}_i$ results in

$$p(\bar{\mathbf{X}} | \mathbf{X}, \boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^I \left(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_x \right)' \boldsymbol{\Omega}_x \left(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_x \right) \right\}, \tag{4.13}$$

where

$$\begin{aligned}
 \boldsymbol{\Omega}_x &= \mathbf{C}^{(x)'} \bar{\mathbf{W}}^{(x)'} \boldsymbol{\Sigma}_x^{-1} \bar{\mathbf{W}}^{(x)} \mathbf{C}^{(x)} + \bar{\mathbf{W}}^{(x)} \mathbf{H} \boldsymbol{\Psi}^{-1} \mathbf{H}' \bar{\mathbf{W}}^{(x)'} + \mathbf{Q}_x^{-1} \\
 \boldsymbol{\mu}_x &= \boldsymbol{\Omega}_x^{-1} \left(\mathbf{C}^{(x)'} \bar{\mathbf{W}}^{(x)'} \boldsymbol{\Sigma}_x^{-1} \mathbf{x}_i + \bar{\mathbf{W}}^{(x)} \mathbf{H} \boldsymbol{\Psi}^{-1} \left(\boldsymbol{\gamma}_i^{(y)} - \mathbf{N} \boldsymbol{\gamma}_i^{(y)} \right) + \mathbf{Q}_x^{-1} \mathbf{m}_x \right)
 \end{aligned} \tag{4.14}$$

Thus, the values of the error-free variables are drawn from the posterior distribution for $\bar{\mathbf{x}}_i$

conditional on the observed data and parameters, which is given by $\bar{\mathbf{x}}_i | \mathbf{x}_i, \boldsymbol{\theta} \sim \text{Normal}(\boldsymbol{\mu}_x, \boldsymbol{\Omega}_x^{-1})$.

The full conditional distribution for $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}_1, \dots, \bar{\mathbf{y}}_I]'$ is obtained as

$$\begin{aligned}
 p(\bar{\mathbf{Y}} | \mathbf{Y}, \boldsymbol{\theta}) &\propto \prod_{i=1}^I p(\mathbf{y}_i | \bar{\mathbf{y}}_i, \boldsymbol{\theta}) p(\bar{\mathbf{y}}_i) \\
 &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^I \left(\mathbf{y}_i - \mathbf{C}^{(y)'} \bar{\mathbf{W}}^{(y)} \bar{\mathbf{y}}_i \right)' \boldsymbol{\Sigma}_y^{-1} \left(\mathbf{y}_i - \mathbf{C}^{(y)'} \bar{\mathbf{W}}^{(y)} \bar{\mathbf{y}}_i \right) + \left(\bar{\mathbf{y}}_i - \mathbf{m}_y \right)' \mathbf{Q}_y^{-1} \left(\bar{\mathbf{y}}_i - \mathbf{m}_y \right) \right\},
 \end{aligned} \tag{4.15}$$

where

$$\begin{aligned} \mathbf{m}_y &= \Delta^{\bar{y}} \left[(\mathbf{I} - \mathbf{S})^{-1} \mathbf{H} \overline{\overline{\mathbf{W}}}^{(x)} \mathbf{m}_x \right] \\ \mathbf{Q}_y^{-1} &= \Delta^{\bar{y}} \left[(\mathbf{I} - \mathbf{S})^{-1} \mathbf{H} \overline{\overline{\mathbf{W}}}^{(x)'} \mathbf{Q}_x \overline{\overline{\mathbf{W}}}^{(x)} \mathbf{H} (\mathbf{I} - \mathbf{N})^{-1} + (\mathbf{I} - \mathbf{N})^{-1} \Psi^{-1} (\mathbf{I} - \mathbf{N})^{-1} \right], \end{aligned} \quad (4.16)$$

and $\Delta^{\bar{y}} = \left(\overline{\overline{\mathbf{W}}}^{(y)} \overline{\overline{\mathbf{W}}}^{(y)'} \right)^{-1} \overline{\overline{\mathbf{W}}}^{(y)}$. Rearranging (4.15) with respect to \bar{y}_i yields

$$p(\bar{\mathbf{Y}} | \mathbf{Y}, \boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^I (\bar{\mathbf{y}}_i - \boldsymbol{\mu}_y)' \boldsymbol{\Omega}_y (\bar{\mathbf{y}}_i - \boldsymbol{\mu}_y) \right\}, \quad (4.17)$$

where $\boldsymbol{\Omega}_y = \mathbf{C}^{(y)'} \overline{\overline{\mathbf{W}}}^{(y)} \boldsymbol{\Sigma}_y^{-1} \overline{\overline{\mathbf{W}}}^{(y)} \mathbf{C}^{(y)} + \mathbf{Q}_y^{-1}$ and $\boldsymbol{\mu}_y = \boldsymbol{\Omega}_y^{-1} \left(\mathbf{C}^{(y)'} \overline{\overline{\mathbf{W}}}^{(y)} \boldsymbol{\Sigma}_y^{-1} \mathbf{y}_i + \mathbf{Q}_y^{-1} \mathbf{m}_y \right)$.

Step 2. This step updates the weights ($\overline{\overline{\mathbf{W}}}$) to obtain the error-free latent variables in (4.2). To obtain the posterior in the same distributional family given the likelihood function in (4.10), the unknown parameters in each column of $\overline{\overline{\mathbf{W}}}^{(x)}$ are assumed to have a normal prior distribution with two following hyperparameters: a J_1 by 1 vector $\mathbf{m}_{w1(p_1)}$ as the mean hyperparameter and a J_1 by J_1 covariance matrix $\mathbf{S}_{w1(p_1)}$ as the covariance hyperparameter. That is,

$\mathbf{w}_{p_1}^{(x)} \sim \text{Normal}(\mathbf{m}_{w1(p_1)}, \mathbf{S}_{w1(p_1)})$ ($p_1 = 1, \dots, P_1$). Combining the prior distributions and the

likelihood function of the observed data in (4.11), the posterior distribution for $\overline{\overline{\mathbf{W}}}^{(x)}$ becomes proportional to

$$\begin{aligned}
 p\left(\overline{\mathbf{W}}^{(x)} \mid \mathbf{X}, \overline{\mathbf{X}}, \boldsymbol{\theta}^{(-\mathbf{W}^{(x)})}\right) &\propto \prod_{i=1}^I \prod_{p_1=1}^{P_1} p\left(\mathbf{x}_i \mid \overline{\mathbf{x}}_i, \mathbf{w}_{p_1}^{(x)}, \mathbf{c}_{p_1}^{(x)}, \boldsymbol{\delta}_i^{(x)}\right) p\left(\mathbf{w}_{p_1}^{(x)}\right) \\
 &\propto \exp\left\{-\frac{1}{2} \sum_{p_1=1}^{P_1} \left(\mathbf{X} - \overline{\mathbf{X}} \mathbf{w}_{p_1}^{(x)} \mathbf{c}_{p_1}^{(x)}\right)' \boldsymbol{\Sigma}_x^{-1} \left(\mathbf{X} - \overline{\mathbf{X}} \mathbf{w}_{p_1}^{(x)} \mathbf{c}_{p_1}^{(x)}\right)\right\} \\
 &\quad \cdot \exp\left\{-\frac{1}{2} \sum_{p_1=1}^{P_1} \left(\mathbf{w}_{p_1}^{(x)} - \mathbf{m}_{\mathbf{w}1(p_1)}\right)' \mathbf{S}_{\mathbf{w}1(p_1)}^{-1} \left(\mathbf{w}_{p_1}^{(x)} - \mathbf{m}_{\mathbf{w}1(p_1)}\right)\right\}, \quad (4.18)
 \end{aligned}$$

where $\boldsymbol{\theta}^{(-\mathbf{W}^{(x)})}$ represents all the other model parameters except $\mathbf{w}_{p_1}^{(x)}$. Consequently, samples from the weights for exogenous latent variables are generated from the following full conditional distribution for $\mathbf{w}_{p_1}^{(x)}$:

$$\mathbf{w}_{p_1}^{(x)} \mid \mathbf{X}, \overline{\mathbf{X}}, \boldsymbol{\theta}^{(-\mathbf{W}^{(x)})} \sim \text{Normal}\left(\boldsymbol{\Phi}_x^{-1} \left(\text{tr}\left(\overline{\mathbf{X}} \mathbf{X}'\right) \boldsymbol{\Sigma}_x^{-1} \mathbf{c}_{p_1}^{(x)'} + \mathbf{S}_{\mathbf{w}1(p_1)}^{-1} \mathbf{m}_{\mathbf{w}1(p_1)}\right), \boldsymbol{\Phi}_x^{-1}\right), \quad (4.19)$$

where $\boldsymbol{\Phi}_x = \mathbf{c}_{p_1}^{(x)'} \mathbf{c}_{p_1}^{(x)} \overline{\mathbf{X}} \overline{\mathbf{X}} \boldsymbol{\Sigma}_x^{-1} + \mathbf{S}_{\mathbf{w}1(p_1)}^{-1}$. Similarly, the posterior distribution for the weights for endogenous latent variables are updated combining a normal prior distribution with

hyperparameters $\mathbf{m}_{\mathbf{w}2(p_2)}$ and $\mathbf{S}_{\mathbf{w}2(p_2)}$, $\mathbf{w}_{p_2}^{(y)} \sim \text{Normal}(\mathbf{m}_{\mathbf{w}2(p_2)}, \mathbf{S}_{\mathbf{w}2(p_2)})$ ($p_2 = 1, \dots, P_2$), and the

likelihood function in (4.11). This results in the following full conditional for $\mathbf{w}_{p_2}^{(y)}$:

$$\mathbf{w}_{p_2}^{(y)} \mid \mathbf{Y}, \overline{\mathbf{Y}}, \boldsymbol{\theta}^{(-\mathbf{W}^{(y)})} \sim \text{Normal}\left(\boldsymbol{\Phi}_y^{-1} \left(\text{tr}\left(\overline{\mathbf{Y}} \mathbf{Y}'\right) \boldsymbol{\Sigma}_y^{-1} \mathbf{c}_{p_2}^{(y)'} + \mathbf{S}_{\mathbf{w}2(p_2)}^{-1} \mathbf{m}_{\mathbf{w}2(p_2)}\right), \boldsymbol{\Phi}_y^{-1}\right), \quad (4.20)$$

where $\boldsymbol{\Phi}_y = \mathbf{c}_{p_2}^{(y)'} \mathbf{c}_{p_2}^{(y)} \overline{\mathbf{Y}} \overline{\mathbf{Y}} \boldsymbol{\Sigma}_y^{-1} + \mathbf{S}_{\mathbf{w}2(p_2)}^{-1}$.

Step 3. This step updates the loadings (\mathbf{C}) in the measurement model. As in *step 2*, a normal

prior distribution for each p_1 th row of $\mathbf{C}^{(x)}$ ($p_1 = 1, \dots, P_1$) is assumed: $\mathbf{c}_{p_1}^{(x)} \sim N(\mathbf{m}_{\mathbf{C}1(p_1)}, \mathbf{S}_{\mathbf{C}1(p_1)})$,

where $\mathbf{m}_{\mathbf{C}1(p_1)}$ is a J_1 by 1 vector of prior means and $\mathbf{S}_{\mathbf{C}1(p_1)}$ is a J_1 by J_1 matrix of prior covariance

matrix. The full conditional for $\mathbf{c}_{p_1}^{(x)}$ follows a normal distribution, that is,

$$\mathbf{c}_{p_1}^{(x)} | \mathbf{X}, \bar{\bar{\mathbf{X}}}, \boldsymbol{\theta}^{(-C^{(x)})} \sim \text{Normal} \left(\boldsymbol{\Pi}_x^{-1} \left(\boldsymbol{\Sigma}_x^{-1} \bar{\bar{\mathbf{X}}} \mathbf{X} \mathbf{w}_{p_1}^{(x)} + \mathbf{S}_{C1(p_1)}^{-1} \mathbf{m}_{C1(p_1)} \right), \boldsymbol{\Pi}_x^{-1} \right), \quad (4.21)$$

where $\boldsymbol{\Pi}_x = \left(\mathbf{w}_{p_1}^{(x)'} \bar{\bar{\mathbf{X}}} \bar{\bar{\mathbf{X}}} \mathbf{w}_{p_1}^{(x)} \right) \boldsymbol{\Sigma}_x^{-1} + \mathbf{S}_{C1(p_1)}^{-1}$. Likewise, samples for the loadings of endogenous

latent variables are drawn from $p \left(\mathbf{c}_{p_2}^{(y)} | \mathbf{y}_i, \boldsymbol{\theta}^{(-C^{(y)})} \right)$. By combining the prior

$\mathbf{c}_{p_2}^{(y)} \sim \text{Normal}(\mathbf{m}_{C2(p_2)}, \mathbf{S}_{C2(p_2)})$ ($p_2 = 1, \dots, P_2$) and the likelihood function in (4.11), the full

conditional for $\mathbf{c}_{p_2}^{(y)}$ is

$$\mathbf{c}_{p_2}^{(y)} | \mathbf{Y}, \bar{\bar{\mathbf{Y}}}, \boldsymbol{\theta}^{(-C^{(y)})} \sim \text{Normal} \left(\boldsymbol{\Pi}_y^{-1} \left(\boldsymbol{\Sigma}_y^{-1} \bar{\bar{\mathbf{Y}}} \mathbf{Y} \mathbf{w}_{p_2}^{(y)} + \mathbf{S}_{C2(p_2)}^{-1} \mathbf{m}_{C2(p_2)} \right), \boldsymbol{\Pi}_y^{-1} \right), \quad (4.22)$$

where $\boldsymbol{\Pi}_y = \left(\mathbf{w}_{p_2}^{(y)'} \bar{\bar{\mathbf{Y}}} \bar{\bar{\mathbf{Y}}} \mathbf{w}_{p_2}^{(y)} \right) \boldsymbol{\Sigma}_y^{-1} + \mathbf{S}_{C2(p_2)}^{-1}$

Step 4. In this step, the variances of the measurement error term are updated. Assuming a

conjugate prior distribution for the variance of each measurement error term, the j_1 th $\phi_{j_1}^{(x)}$ ($j_1 =$

$1, \dots, J_1$) follows an inverse gamma distribution with a shape hyperparameter $a_{x0(j_1)}$ and scale

hyperparameter $\nu_{x0(j_1)}$. That is, $\phi_{j_1}^{(x)} \sim \text{IG}(a_{x0(j_1)}, \nu_{x0(j_1)})$. The full conditional for $\phi_{j_1}^{(x)}$ is written as:

$$\begin{aligned} p \left(\phi_{j_1}^{(x)} | \mathbf{x}_i, \boldsymbol{\theta}^{(-\phi^{(x)})} \right) &\propto \phi_{j_1}^{(x) \frac{-I}{2}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{J_1} \phi_{j_1}^{(x)-1} \sum_{i=1}^I \left(x_{i(j_1)} - \bar{\gamma}_i^{(x)} \mathbf{C}_j^{(x)} \right)^2 \right\} \phi_{j_1}^{(x) - a_{x0(j_1)} - 1} \exp \left\{ -\phi_{j_1}^{(x)-1} \nu_{x0(j_1)} \right\} \\ &= \phi_{j_1}^{(x) - \left(\frac{I}{2} + a_{x0(j_1)} + 1 \right)} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{J_1} \phi_{j_1}^{(x)-1} \left[\nu_{x0(j_1)} + \sum_{i=1}^I \left(x_{i(j_1)} - \bar{\gamma}_i^{(x)} \mathbf{C}_j^{(x)} \right)^2 \right] \right\}. \end{aligned} \quad (4.23)$$

This shows that the full conditional distribution for $\phi_{j_1}^{(x)}$ given the observed variable $x_{i(j_1)}$ and

other parameters is given by

$$\phi_{j_1}^{(x)} | \mathbf{x}_i, \boldsymbol{\theta}^{(-\phi^{(x)})} \sim \text{IG} \left(\frac{I}{2} + a_{x0(j_1)}, \nu_{x0(j_1)} + \frac{1}{2} \sum_{i=1}^I \left(x_{i(j_1)} - \bar{\gamma}_i^{(x)} \mathbf{C}_j^{(x)} \right)^2 \right). \quad (4.24)$$

Similarly, the posterior distributions for the variances of $\delta_i^{(y)}$ are updated with the same inverse gamma distribution but with different shape and scale hyperparameters $a_{Y0(j_2)}$ and $\nu_{Y0(j_2)}$, respectively ($j_2 = 1, \dots, J_2$). The full conditional for $\phi_{j_2}^{(y)}$ follows an inverse gamma distribution,

$$\phi_{j_2}^{(y)} | y_i, \boldsymbol{\theta}^{(-\phi^{(y)})} \sim \text{IG} \left(\frac{I}{2} + a_{Y0(j_2)}, \nu_{Y0(j_2)} + \frac{1}{2} \sum_{i=1}^I \left(x_{i(j_2)} - \boldsymbol{\gamma}_i^{(y)} \mathbf{C}_{j_2}^{(y)} \right)^2 \right) \quad (4.25)$$

Step 5. This step updates the path coefficients (\mathbf{V}) in the structural model. From samples generated in step 1 and step 2, the error-free latent variables are imputed according to the

weighted relation model in (4.2). Let $\overline{\boldsymbol{\Gamma}}^{(x)} = \left[\overline{\boldsymbol{\gamma}}_1^{(x)}, \dots, \overline{\boldsymbol{\gamma}}_I^{(x)} \right]'$ denote an I by P_1 matrix of error-free

exogenous latent variables, and $\overline{\boldsymbol{\Gamma}}^{(y)} = \left[\overline{\boldsymbol{\gamma}}_1^{(y)}, \dots, \overline{\boldsymbol{\gamma}}_I^{(y)} \right]'$ denote an I by P_2 matrix of error-free

endogenous latent variables. Using (4.4) and (4.9), the likelihood function of the endogenous

latent variables is written as: $p \left(\overline{\boldsymbol{\Gamma}}^{(y)} | \overline{\boldsymbol{\Gamma}}^{(x)}, \boldsymbol{\theta} \right) = \prod_{i=1}^I \text{Normal} \left(\overline{\boldsymbol{\Gamma}}_i \mathbf{V}', \boldsymbol{\Psi} \right)$. Given $\overline{\boldsymbol{\gamma}}_i^{(x)}$, each endogenous

latent variable is assumed to be independent of each other, and so that the p_2 th latent variables

can be updated separately. Let $\mathbf{V}_{(p_2)}^\oplus$ denote a row vector consisting of the unknown parameters

only in the p_2 th row of \mathbf{V} . Let $\overline{\boldsymbol{\Gamma}}^\oplus = \left[\overline{\boldsymbol{\Gamma}}^{(x)} \quad \overline{\boldsymbol{\Gamma}}^{(y)} \right]$. Then, $\overline{\boldsymbol{\Gamma}}_{(\sim p_2)}^\oplus$ is a matrix of columns of $\overline{\boldsymbol{\Gamma}}^\oplus$

corresponding to non-zero elements in the p_2 th row of \mathbf{V} . The prior distribution for $\mathbf{V}_{(p_2)}^\oplus$ is

specified to follow a normal distribution with hyperparameters of $\mathbf{m}_{V(p_2)}$ and $\mathbf{S}_{V(p_2)}$:

$\mathbf{V}_{(p_2)}^{\oplus'} \sim \text{Normal} \left(\mathbf{m}_{V(p_2)}, \mathbf{S}_{V(p_2)} \right)$. The posterior distribution for $\mathbf{V}_{(p_2)}^\oplus$ becomes proportional to

$p \left(\overline{\boldsymbol{\Gamma}}_{p_2}^{(y)} | \overline{\boldsymbol{\Gamma}}^{(x)}, \boldsymbol{\theta} \right) p \left(\mathbf{V}_{(p_2)}^\oplus \right)$, where $\overline{\boldsymbol{\Gamma}}_{p_2}^{(y)}$ refers to an I by 1 vector of the p_2 th latent variable scores.

Hence, the samples of unknown parameters in \mathbf{V} are drawn from the full conditional for $\mathbf{V}_{(p_2)}^\oplus$, which is given by

$$\mathbf{V}_{(p_2)}^\oplus \mid \bar{\mathbf{X}}, \bar{\mathbf{Y}}, \boldsymbol{\theta}^{(-\mathbf{V})} \sim \text{Normal} \left(\boldsymbol{\Omega}_{\mathbf{V}}^{-1} \left(\phi_{p_2}^{(\gamma)^{-1}} \begin{pmatrix} \bar{\boldsymbol{\Gamma}}_{(\sim p_2)}^\oplus & \bar{\boldsymbol{\Gamma}}_{p_2}^{\prime(\mathbf{y})} \end{pmatrix} + \mathbf{S}_{\mathbf{V}(p_2)}^{-1} \mathbf{m}_{\mathbf{V}(p_2)} \right), \boldsymbol{\Omega}_{\mathbf{V}}^{-1} \right), \quad (4.26)$$

where $\boldsymbol{\Omega}_{\mathbf{V}} = \begin{pmatrix} \bar{\boldsymbol{\Gamma}}_{(\sim p_2)}^\oplus & \bar{\boldsymbol{\Gamma}}_{(\sim p_2)}^{\prime(\mathbf{y})} \\ \bar{\boldsymbol{\Gamma}}_{(\sim p_2)}^{\prime(\mathbf{y})} & \bar{\boldsymbol{\Gamma}}_{p_2}^{\prime(\mathbf{y})} \end{pmatrix} \phi_{p_2}^{(\gamma)^{-1}} + \mathbf{S}_{\mathbf{V}(p_2)}^{-1}$.

Step 6. This step updates the variances of the structural errors ζ_i (i.e., $\phi_1^{(\gamma)}, \dots, \phi_{p_2}^{(\gamma)}$), given the same likelihood function used in the previous step. As used in the measurement error terms, the prior distribution for each structural error is specified as an inverse gamma distribution with a shape and scale hyperparameters, $a_{\text{LV0}(p_2)}$ and $\nu_{\text{LV0}(p_2)}$, respectively. The full conditional

distribution for $\phi_{p_2}^{(\gamma)}$ follows an inverse gamma distribution with a shape parameter of $\frac{I}{2} + a_{\text{LV0}(p_2)}$

and a scale parameter of $\nu_{\text{LV0}(p_2)} + \frac{1}{2} \sum_{i=1}^I \left(\gamma_{i(p_2)}^{(\mathbf{y})} - \mathbf{V}_{(p_2)}^\oplus \bar{\boldsymbol{\Gamma}}_{i(\sim p_2)}^\oplus \right)^2$.

In BGSCA, the parameters of the prior distributions, called prior hyperparameters, are assumed to be known constants. Their values are chosen based on our belief or prior knowledge about the magnitude of $\boldsymbol{\theta}$. While a mean hyperparameter is to specify the central location of a prior distribution, the variance hyperparameter determines the variability or magnitude of the spread of a prior. If we have very little prior knowledge about what value a parameter might take, then we may adopt a prior that considers a wide range of values of the parameter, involving a large variance hyperparameter value. Such a prior is called a diffuse or uninformative prior. For example, a normal prior distribution with a variance hyperparameter value of 100 amounts to a

diffuse prior. On the other hand, when we have substantial prior knowledge about the variance of the parameter, we would specify a prior distribution with a small variance value.

We also need to set the number of total iterations (S) as well as burn-in sample size. Typically, the number of total iterations is set as 10,000, whose first 1000 iterations are discarded as the burn-in sample size. Convergence of the MCMC chain is evaluated via drawing trace plots. If a trace plot displays some trends other than shifting up and down (e.g., staying in a certain values long), it indicates that there is no convergence reached yet. In such cases, albeit computationally intensive, we would need to increase S , and/or discard a bigger proportion of early iterations.

4.3. Other Extensions

This section discusses other possible technical extensions that would further improve the generality of BGSCA. These extensions pertain to the accommodation of correlated error terms in the measurement model (Section 4.3.1) and the analysis of categorical data (Section 4.3.2).

4.3.1. *Correlated error terms in measurement models*

The estimation procedure in Section 4.2 was developed under the assumption that each error term in the measurement model is independent of one another, yielding a diagonal covariance matrix for the measurement error terms. As stated earlier, although this is a conventional assumption made in standard linear SEMs, it is not required and can be relaxed. Practically, we may hypothesize that two or more observed variables share something in common, which is not explained by their latent variables explicitly specified in the model. To accommodate such situations, BGSCA can be extended to relax independence of the observed variables by allowing the estimation of both variances and covariances of the measurement error terms. This can be technically done by modifying the specification of a conjugate prior

distribution for the covariance matrix of the measurement error terms as an inverse Wishart distribution. Specifically, *step 4* in Section 4.2 is changed as follows:

Step 4. A covariance matrix of the measurement error term Σ_x is updated by using a conjugate prior distribution, an inverse Wishart distribution. The inverse Wishart distribution for a J_1 by J_1 matrix Σ_x is expressed as: $\Sigma_x \sim \text{IW}(\mathbf{R}_{x_0}, \nu_{x_0})$, where \mathbf{R}_{x_0} is a J_1 by J_1 positive definite matrix and ν_{x_0} a positive integer. The inverse Wishart distribution is considered a multivariate extension of an inverse gamma distribution. The hyperparameters of an inverse Wishart prior distribution are often set as results obtained via maximum likelihood (e.g., Bouriga & Féron, 2013; Haff, 1980), and such method is called an empirical Bayes. Empirical Bayes is a procedure of estimating a prior distribution from the data. By combining the specified inverse Wishart prior with the likelihood function of data, the full conditional distribution for Σ_x is expressed as:

$$p\left(\Sigma_x \mid \bar{\mathbf{X}}, \boldsymbol{\theta}^{(-\Sigma_x)}\right) \propto |\Sigma_x|^{-\frac{I}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^I \left(\mathbf{x}_i - \bar{\boldsymbol{\gamma}}_i \mathbf{C}^{(x)}\right)' \Sigma_x^{-1} \left(\mathbf{x}_i - \bar{\boldsymbol{\gamma}}_i \mathbf{C}^{(x)}\right)\right\} \\ \times |\Sigma_x|^{-\frac{(\nu_{x_0} + J_1 + 1)}{2}} \exp\left\{-\frac{1}{2} \text{tr}[\Sigma_x^{-1} \mathbf{R}_{x_0}]\right\}. \quad (4.27)$$

A covariance matrix of the measurement errors for \mathbf{x}_i is sampled from the full conditional for Σ_x , that is,

$$\Sigma_x \mid \bar{\mathbf{X}}, \boldsymbol{\theta}^{(-\Sigma_x)} \sim \text{IW}\left(\sum_{i=1}^I \left(\mathbf{x}_i - \bar{\boldsymbol{\gamma}}_i \mathbf{C}^{(x)}\right)' \left(\mathbf{x}_i - \bar{\boldsymbol{\gamma}}_i \mathbf{C}^{(x)}\right) + \mathbf{R}_{x_0}, I + \nu_{x_0}\right). \quad (4.28)$$

Likewise, the posterior distribution for the covariances and variances of the endogenous observed variables Σ_y is updated by combining the data likelihood function and an inverse Wishart prior distribution with two hyperparameters \mathbf{R}_{y_0} and ν_{y_0} . The full conditional for Σ_y is

$\Sigma_y | \bar{\mathbf{Y}}, \theta^{(-\Sigma_y)} \sim \text{IW} \left(\sum_{i=1}^I \left(\mathbf{y}_i - \gamma_i^{(y)} \mathbf{C}^{(y)} \right)' \left(\mathbf{y}_i - \gamma_i^{(y)} \mathbf{C}^{(y)} \right) + \mathbf{R}_{y_0}, I + \nu_{y_0} \right)$, and samples for the

covariances and variances of the measurement errors are generated accordingly.

4.3.2. Analysis of categorical data

Until now, the observed data are assumed to be continuous and sampled from a multivariate normal distribution. However, data in the social sciences are often categorical in nature. In particular, perhaps the most common type of categorical data in SEMs is ordered categorical, e.g., questionnaire items on three-, five-, or seven- point Likert scales (e.g., Lee, 2012; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Because such categorical data would likely follow skewed distributions, treating them simply as continuous and normal would likely yield misleading results, including biased parameter estimates (Muthén & Kaplan, 1985; Rhemtulla et al., 2012).

BGSCA is extended to analyze categorical data based on the latent response variable formulation, as in a Bayesian approach to factor-based SEM (e.g., Lee, 2012, Chapter 5). The basic idea of the latent response variable formulation is that responses of each categorical variable come from an underlying continuous variable that follows a normal distribution. A set of thresholds is then used to discretize the normal distribution, resulting in non-equidistance intervals of the observed categories (e.g., Finney & DiStefano, 2006; Lee & Song, 2004). Suppose that \mathbf{z}_i now contains (J_1+J_2) categorical variables. Let L denote the number of response choices of each categorical variable \mathbf{z}_{ik} ($k = 1, \dots, J_1+J_2$). With the latent response variable formulation, the observed response \mathbf{z}_{ik} is thought to arise from a latent response variable \mathbf{z}_{ik}^* as follows:

$$\mathbf{z}_{ik} = \begin{cases} 1 & \text{if } \mathbf{z}_{ik}^* \leq \alpha_1 \\ 2 & \text{if } \alpha_1 < \mathbf{z}_{ik}^* \leq \alpha_2 \\ \vdots & \vdots \\ L-1 & \text{if } \alpha_{L-2} < \mathbf{z}_{ik}^* \leq \alpha_{L-1} \\ L & \text{if } \alpha_{L-1} < \mathbf{z}_{ik}^* \end{cases}, \quad (4.29)$$

where α_l is a category threshold discretizing the underlying continuous response variable into observed categorical responses ($l=1, 2, \dots, L-1$). Generally speaking, within this formation, the category thresholds linking the observed responses and underlying continuous latent responses are first updated. The \mathbf{z}_{ik} in (4.3) is in turn replaced with the estimated \mathbf{z}_{ik}^* in the measurement model for the subsequent steps for parameter estimation in BGSCA.

To implement the latent response variable formulation, a Metropolis-Hastings algorithm is included as an additional step (e.g., Lee, 2012) before applying the proposed BGSCA algorithm. Let $\boldsymbol{\alpha}_k$ denote a 1 by $(L-1)$ vector of category thresholds for the k th observed variable, $\boldsymbol{\alpha}_k = [\alpha_{1k}, \dots, \alpha_{(L-1)k}]$. From a Metropolis-Hastings algorithm, at the s th iteration, possible new candidate values for $\boldsymbol{\alpha}_k$ are simulated from a truncated normal proposal distribution defined on a bounded range $[\alpha_{(j-1)k}, \alpha_{(j+1)k}]$ ($j = 1, \dots, L-1$). Because threshold values in $\boldsymbol{\alpha}_k$ define a set of specific ranges that each category would belong in a continuum scale, a truncated normal distribution seems appropriate. Then, $\boldsymbol{\alpha}_k^{(s)}$ ($s = 1, \dots, S$) are successively generated according to the following algorithm.

Step 1. Set $s = 1$, and generate a vector of new candidate values $\boldsymbol{\alpha}_k^*$ at random from a

proposal density function $q(\boldsymbol{\alpha}_k^* | \boldsymbol{\alpha}_k^{(s-1)})$, where $\boldsymbol{\alpha}_k^{(0)}$ is a vector of randomly chosen initial values.

Step 2. Given $\boldsymbol{\alpha}_k^*$, compute the acceptance ratio between the candidate ($\boldsymbol{\alpha}_k^*$) and current

$(\boldsymbol{\alpha}_k^{(s-1)})$ values as

$$R = \frac{p(\boldsymbol{\alpha}_k^*, \mathbf{z}_{ik}^* | \mathbf{z}_{ik}, \bar{\boldsymbol{\Gamma}}_i, \boldsymbol{\theta}) q(\boldsymbol{\alpha}_k^{(s-1)} | \boldsymbol{\alpha}_k^*)}{p(\boldsymbol{\alpha}_k^{(s-1)}, \mathbf{z}_{ik}^* | \mathbf{z}_{ik}, \bar{\boldsymbol{\Gamma}}_i, \boldsymbol{\theta}) q(\boldsymbol{\alpha}_k^* | \boldsymbol{\alpha}_k^{(s-1)})} \quad (4.30)$$

Step 3. Decide whether to accept the candidate vector $\boldsymbol{\alpha}_k^*$ with probability of $\min\{1, R\}$.

Keep the current values $\boldsymbol{\alpha}_k^{(s-1)}$ otherwise.

Step 4. For the accepted candidate value only, simulate a new set of \mathbf{z}_{ik}^* from a truncated normal distribution.

Step 5. Set $s = s + 1$, and repeat steps 2 - 4 until $s = S$.

Chapter 5. Simulation Study

This chapter presents two simulation studies. The first simulation study is conducted to evaluate the performance of BGSCA, focusing on how the amount of variability in the error-free data, the magnitude of measurement error variances, and sample size would affect the parameter recovery capability of BGSCA. The second simulation is to study the effects of specifying different hyperparameters for the prior distributions on the performance of BGSCA. This can provide a general guideline on how to choose hyperparameter values for BGSCA.

5.1. Simulation Study 1

5.1.1. Simulation Design and Data Generation

In the first simulation study, we considered a model that involved three latent variables, each of which consisted of five observed variables. The structural model was hypothesized such that the first two error-free latent variables, $\bar{\gamma}_1$ and $\bar{\gamma}_2$, influenced the other error-free latent variable, $\bar{\gamma}_3$, as shown in Figure 5- 1. This led to $P_1 = 2$, $P_2 = 1$, $J_1 = 10$, and $J_2 = 5$. All free parameter values in the weight matrix ($\bar{\bar{W}}$) were prescribed to 0.3, and those in the loading matrix (\mathbf{C}) were to 0.8. We also chose two path coefficients in \mathbf{V} (b_1 and b_2) and the error variance of the endogenous latent variable ($\text{var}(\zeta)$) as 0.6, 0.4, and 0.48, respectively.

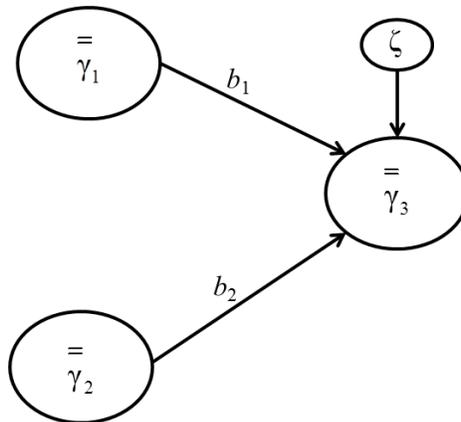


Figure 5- 1. The structural model specified for simulation study 1.

There were three manipulated factors in this simulation study. The first factor was the amount of the variance of the error-free observed exogenous variables ($\text{var}(\mathbf{Q}_x)$) for generating the error-free latent variables, and it varied at three levels: 1, 2, and 3. The second factor was the magnitude of the measurement error variances ($\text{var}(\delta_i) = \phi$), which was also manipulated to vary at three levels: 0.3, 1, and 2. The last factor was the number of subjects (I), varying at three levels: 50, 100, and 200. Thus, this study involved $3 \times 3 \times 3 = 27$ conditions. For each condition, 500 simulated sets were generated, thus yielding a total of 13,500 data sets.

To generate a data set, a P_1 by 1 vector of the exogenous latent variable scores for the i th subject $\overline{\gamma}_i^{(x)} = \begin{bmatrix} \overline{\gamma}_{i1} \\ \overline{\gamma}_{i2} \end{bmatrix}$ was first generated. Let \mathbf{m}_x denote an J_1 by 1 mean vector of zeros, and \mathbf{Q}_x

denote an J_1 by J_1 diagonal matrix whose diagonal element was one of the three levels of $\text{var}(\mathbf{Q}_x)$. Let $\overline{\mathbf{x}}_i$ denote a J_1 by 1 vector of error-free observed exogenous variables. Assuming

$\overline{\mathbf{x}}_i \sim \text{Normal}(\mathbf{m}_x, \mathbf{Q}_x)$, $\overline{\gamma}_i^{(x)}$ was generated from $\text{Normal}\left(\mathbf{m}_x \overline{\mathbf{W}}^{(x)}, \overline{\mathbf{W}}^{(x)'} \mathbf{Q}_x \overline{\mathbf{W}}^{(x)}\right)$. This way of

generating these latent variables was derived based on the weighted relation model of BGSCA

defined in (4.2). Then, the endogenous latent variable $\overline{\gamma}_{i3}$ was generated from the structural

model as specified in Figure 5- 1. The error term ζ in the structural model was assumed to be

uncorrelated with the two exogenous latent variables (i.e., $\text{cov}(\overline{\gamma}_1, \zeta) = \text{cov}(\overline{\gamma}_2, \zeta) = 0$), and thus

the variance of the endogenous latent variable was set as

$\text{var}(\overline{\gamma}_3) = b_1^2 \text{var}(\overline{\gamma}_1) + b_2^2 \text{var}(\overline{\gamma}_2) + 2b_1b_2 \text{cov}(\overline{\gamma}_1, \overline{\gamma}_2) + \text{var}(\zeta)$. Note that assuming the diagonal

matrix for \mathbf{Q}_x set the third term of $\text{var}(\overline{\gamma}_3)$ to be equal to zero. Given the latent variables

generated, their corresponding observed variables were subsequently generated based on the measurement model defined in (4.3).

Each of the 13,500 data sets was analyzed by BGSCA with diffuse prior distributions, which were the most commonly used specification in Bayesian analyses (e.g., Spiegelhalter, Thomas, Best, & Gilk, 2003; Chen, Bakshi, & Goel, 2009). Specifically, for $\bar{\mathbf{z}}_i$, $\bar{\mathbf{W}}$, \mathbf{C} , and \mathbf{V} , normal distributions were assumed with mean hyperparameter value of zero and variance hyperparameter of 100. To specify diffuse prior distributions for $\phi_1^{(x)}, \dots, \phi_{J_1}^{(x)}$, $\phi_1^{(y)}, \dots, \phi_{J_2}^{(y)}$, and $\text{var}(\zeta)$, inverse-gamma distributions with both shape and scale hyperparameter values of 0.0001 were assumed, each of which would resemble a uniform distribution. A MATLAB code was written to update the posterior distributions for the parameters in BGSCA.

5.1.2. Results

To evaluate parameter recovery, the mean square errors (MSE) of the parameter estimates were calculated as $\text{MSE} = \mathbb{E}[(\omega_{est} - \theta)^2]$, where θ and ω_{est} denote a parameter and its estimate – here, the posterior mean. Although there is no clear-cut criterion to be considered an acceptable level of MSE, it should be a small value close to zero. Table 5- 1 provides the average MSE values of the weight estimates. As shown in the table, when the sample size increased, the MSE values tended to decrease regardless of the amount of variability in the error-free data and the magnitude of the measurement error variances. When $I \geq 100$, the MSE values of the weight estimates were very small, i.e., less than 0.01, except for the cases where the measurement error variances were large. At each sample size, moreover, the changes of the MSE values over the different values of ϕ were negligible. Thus, overall, BGSCA recovered the weights well under the different conditions.

Table 5- 1. The average values of mean square error (MSE) of the estimates in the weighted relation model obtained from BGSCA.

	N	$\text{var}(\mathbf{Q}_x) = 1$			$\text{var}(\mathbf{Q}_x) = 2$			$\text{var}(\mathbf{Q}_x) = 3$		
		$\phi = 0.3$	$\phi = 1$	$\phi = 2$	$\phi = 0.3$	$\phi = 1$	$\phi = 2$	$\phi = 0.3$	$\phi = 1$	$\phi = 2$
$\overline{\mathbf{W}}^{(x)}$	50	0.0430	0.0459	0.0549	0.0442	0.0453	0.0457	0.0449	0.0457	0.0530
	100	0.0041	0.0047	0.0051	0.0042	0.0049	0.0052	0.0041	0.0044	0.0052
	200	0.0019	0.0021	0.0048	0.0017	0.0021	0.0046	0.0019	0.0020	0.0049
$\overline{\mathbf{W}}^{(y)}$	50	0.0361	0.0438	0.0707	0.0362	0.0420	0.0701	0.0363	0.0420	0.0719
	100	0.0042	0.0085	0.0128	0.0041	0.0082	0.0129	0.0046	0.0085	0.0133
	200	0.0031	0.0011	0.0041	0.0032	0.0017	0.0047	0.0031	0.0018	0.0041

Table 5- 2 presents the average MSE values of the loadings and error term estimates in the measurement model. The MSE values of both sets of the estimates decreased with the sample size, regardless of the levels of the other factors. The MSE values also remained similar across the different levels of $\text{var}(\mathbf{Q}_x)$, suggesting that the performance of BGSCA was reasonably stable against changes in the amount of variability in the error-free data. Nevertheless, at each level of $\text{var}(\mathbf{Q}_x)$ and when $I = 50$, the MSE values increased rather sharply as ϕ increased. This may be attributed to the large amount of uncertainties in both sources of information (i.e., information from the prior distributions and the observed data). When less information was given from the observed data with the small sample size, the resulting posterior distributions would be more influenced by the information provided from the prior distributions. However, specifying diffuse priors indicated that there was little information to suggest which values of the parameter might be more plausible than others. Thus, when $I = 50$, insufficient information on both sources of information might contribute to the steep increase in the MSE values. As the sample size increased, such dramatic changes in the MSE values did not occur, being complemented by more information provided from the observed data.

Table 5- 2. The average values of MSE of the estimates in the measurement and structural model obtained from BGSCA.

	I	$\text{var}(\mathbf{Q}_x) = 1$			$\text{var}(\mathbf{Q}_x) = 2$			$\text{var}(\mathbf{Q}_x) = 3$		
		$\phi = 0.3$	$\phi = 1$	$\phi = 2$	$\phi = 0.3$	$\phi = 1$	2	$\phi = 0.3$	$\phi = 1$	$\phi = 2$
$\mathbf{C}^{(x)}$	50	0.0014	0.0204	0.0955	0.0014	0.0207	0.0958	0.0014	0.0205	0.0957
	100	0.0007	0.0173	0.0178	0.0010	0.0172	0.0178	0.0017	0.0171	0.0187
	200	0.0006	0.0148	0.0175	0.0006	0.0144	0.0180	0.0003	0.0148	0.0187
$\mathbf{C}^{(y)}$	50	0.0025	0.0436	0.0888	0.0028	0.0420	0.0881	0.0026	0.0429	0.0869
	100	0.0012	0.0357	0.0387	0.0011	0.0359	0.0382	0.0014	0.0365	0.0387
	200	0.0011	0.0020	0.0073	0.0012	0.0016	0.0080	0.0007	0.0018	0.0084
$\Sigma^{(x)}$	50	0.0254	0.0892	0.2169	0.0262	0.0858	0.2036	0.0257	0.0897	0.2269
	100	0.0047	0.0326	0.0633	0.0063	0.0303	0.0628	0.0053	0.0269	0.0655
	200	0.0028	0.0030	0.0094	0.0035	0.0036	0.0098	0.0031	0.0036	0.0107
$\Sigma^{(y)}$	50	0.0110	0.0606	0.2466	0.0136	0.0628	0.2122	0.0143	0.0678	0.2140
	100	0.0053	0.0369	0.0718	0.0073	0.0362	0.0729	0.0072	0.0389	0.0717
	200	0.0026	0.0072	0.0154	0.0036	0.0086	0.0143	0.0030	0.0098	0.0157

Table 5- 3 shows the average values of MSE of the path coefficient and error variance estimates in the structural model. The general trend of the MSE values was similar to that shown in Table 5- 2, except that the rate of increase in the MSE values across the different levels of ϕ seemed to be less dramatic. When the sample size was large ($I = 200$), all MSE values under the different conditions were smaller than 0.01. In sum, in the weighted relation and structural model, BGSCA was able to recover the parameters sufficiently well across all conditions. In the measurement model, BGSCA also performed well unless the sample size was too small ($I \leq 50$) and at the same time the measurement error variance was too large ($\phi \geq 2$).

Table 5- 3. The average values of MSE of the estimates in the structural model obtained from BGSCA.

	I	$\text{var}(\mathbf{Q}_x) = 1$			$\text{var}(\mathbf{Q}_x) = 2$			$\text{var}(\mathbf{Q}_x) = 3$		
		$\phi = 0.3$	$\phi = 1$	$\phi = 2$	$\phi = 0.3$	$\phi = 1$	$\phi = 2$	$\phi = 0.3$	$\phi = 1$	$\phi = 2$
\mathbf{V}	50	0.0106	0.0350	0.0745	0.0110	0.0335	0.0743	0.0111	0.0343	0.0739
	100	0.0074	0.0130	0.0408	0.0076	0.0126	0.0456	0.0083	0.0122	0.0465
	200	0.0031	0.0060	0.0147	0.0031	0.0059	0.0137	0.0034	0.0053	0.0138
Ψ	50	0.0186	0.0244	0.0851	0.0211	0.0245	0.0872	0.0195	0.0217	0.0889
	100	0.0139	0.0235	0.0683	0.0142	0.0211	0.0667	0.0152	0.0214	0.0664
	200	0.0050	0.0064	0.0150	0.0042	0.0066	0.0152	0.0039	0.0072	0.0154

5.2. Simulation Study 2

5.2.1. Simulation Design

The second simulation study focused on examining how robust BGSCA results are to different specifications of prior distributions. This is called a sensitivity analysis in the Bayesian framework (e.g., Berger, 1990; Depaoli, 2014; Ghosh, Delampady, & Samanta, 2007). As weights ($\overline{\mathbf{W}}$), loadings (\mathbf{C}), and path coefficients (\mathbf{V}) were often the main parameters of interest in SEM, we focused on investigating the sensitivity of their posterior results. For this second study, we used a subset of the simulated data generated from the previous study under $\text{var}(\mathbf{Q}_x) = 1, \phi = 1$, and $I = 100$, and considered four different conditions while varying the mean and variance hyperparameters of the prior distributions for $\overline{\mathbf{W}}, \mathbf{C}, \mathbf{V}$ as listed in Table 5- 4.

Table 5- 4. Specification of different hyperparameter values.

	Hyperparameter values for the priors					
	Weights		Loadings		Path coefficients	
	mean	variance	Mean	variance	mean	variance
Condition 1	0.4	0.1	0.7	0.1	0.5	0.1
Condition 2	0.4	100	0.7	100	0.5	100
Condition 3	0.8	0.1	0.0	0.1	0.0	0.1
Condition 4	0.8	100	0.0	100	0.0	100

In the table, the first condition was regarded an accurate and informative prior specification, because the mean hyperparameters were set to be close to the prescribed true values and their variance hyperparameters were all specified with a small value (being certain about the corresponding mean hyperparameters). The second condition was specified as an accurate but diffuse (or uninformative) prior distribution with a larger variance hyperparameter value. In contrast to the first two conditions, the third and fourth conditions were considered inaccurate prior specifications because their mean hyperparameter values were set to be far from the corresponding prescribed true values. Expressing the extent of uncertainty about the mean hyperparameters, the third condition was an informative prior with a small variance hyperparameter, whereas the fourth condition was considered a diffuse prior with a large variance value. For each condition, 500 simulated data sets were analyzed by BGSCA. The prior distributions for $\bar{\mathbf{z}}_i$, $\phi_1^{(x)}, \dots, \phi_{J_1}^{(x)}$, $\phi_1^{(y)}, \dots, \phi_{J_2}^{(y)}$, and $\text{var}(\zeta)$ remained the same as the previous simulation study. The same MATLAB code was used again for this study.

5.2.2. Results and Discussion

As in the previous study, the mean square errors (MSE) of the estimates of $\bar{\mathbf{W}}$, \mathbf{C} , \mathbf{V} were calculated as a measure of parameter recovery. Table 5- 5 provides the average MSE values of the parameter estimates across the four different specifications of prior distributions. In the table, all MSE values from the first condition were always smaller than those from the third condition. Similarly, the MSE values from the second condition were smaller than those from the fourth condition. It indicated that given the same variability, as expected, an accurate prior specification led to a good approximation to the true values. When comparing the MSE values regardless of the variance hyperparameters, a relatively good approximation to the true values was also achieved with their mean hyperparameters being accurately specified. Importantly, even in

the cases of specifying the mean hyperparameter inaccurately, the MSE values tended to be close to zero if it was paired with a large variance hyperparameter. Consequently, BGSCA was generally robust to the misspecification of prior specifications as long as their variance hyperparameters were set to be large.

Table 5- 5. The average values of MSE of the estimates of weights, loadings, and path coefficients across four different specifications of hyperparameter values.

Parameters	Condition1	Condition2	Condition3	Condition4
$\bar{\mathbf{w}}^{(x)}$	0.0035	0.0036	0.0258	0.0083
$\bar{\mathbf{w}}^{(y)}$	0.0059	0.0062	0.0199	0.0097
$\mathbf{C}^{(x)}$	0.0024	0.0045	0.0239	0.0167
$\mathbf{C}^{(y)}$	0.0017	0.0037	0.0403	0.0349
\mathbf{V}	0.0054	0.0075	0.0213	0.0098

Although this simulation study was useful to examine the performance of BGSCA under the misspecification of prior distributions, it can be further elaborated. For example, the study only manipulated the values of the hyperparameters, while holding all the other factors, such as the sample size or the magnitude of the measurement error variance, constant. It would be worthwhile to assess BGSCA’s robustness to the misspecification, considering the other factors at the same time. In addition, the study was designed to assign user-specific, known values to the hyperparameters. However, these values may be decided in two other ways. One way is to treat the hyperparameters themselves as another set of unknown parameters, so that they also involve their own prior distributions, called hyper-priors, as another level of uncertainty in the modeling process (e.g., Gelman, 2004; Gilks et al, 1996). The other is to apply an Empirical Bayes method, in which the hyperparameter values are determined from the observed data (e.g., Carlin & Louis, 2000).

Chapter 6. An Empirical Example

6.1. Data Description and Model Specification

The present example is a subset of the organizational identification survey data used in Bergami and Bagozzi (2000) (available at <http://www.sem-gesca.org/>). In this example, 305 employees (male = 157 and female = 148) from the electronics division of a large conglomerate in South Korea were asked to evaluate how strongly they agree on 21 survey items or observed variables associated with four constructs or latent variables, called *organizational prestige*, *organizational identification*, *affective commitment (joy)* and *affective commitment (love)*.

Organizational prestige represented an employee's beliefs about whether her/his significant others would highly regard one's organization. *Organizational identification* was capturing how one perceives about the company and oneself within the company. *Affective commitment (joy)* was a type of emotional attachment to the organization, and hence higher latent scores of *joy* was an indicative of "happiness arisen from the organization" (Bergami & Bagozzi, 2000, p. 560). Compared to *joy*, *affective commitment (love)* involved somewhat stronger emotional states, defined as "emotional attraction or affection towards the organization" (Bergami & Bagozzi, 2000, p. 560).

Each latent variable in the model was assumed to underlie a set of observed variables, as shown in Table 6- 1. In the table, *Organizational prestige* (OP) and *organizational identification* (OI) were assumed to be associated with eight (op1 – op8) and six (oi1 – oi6) observed variables, respectively, measured using Mael and Ashforth's (1992) scale. *Affective commitment (joy)* and *affective commitment (love)* were assumed to be associated with four (acj1 – acj4) and three (acl1 – acl3) observed variables, respectively, whose scales were developed by Allen and Meyer (1990). Note that the three observed variables for *affective commitment (love)* were negatively

worded: an employee with a lower score on each of these variables was more likely to feel stronger affection toward the organization. The relationships among the four latent variables were hypothesized in the structural model as displayed in Figure 6- 1. It was assumed that *organizational prestige* affected *organizational identification*, which in turn influenced both types of *affective commitments*.

Table 6- 1. Twenty-one observed variables for four latent variables in Bergami and Bagozzi's (2000) organizational identification data.

Latent variables ¹	Observed variables	
	Label	Description
OP	op1	My relatives and people close or important to me believe that [Company X] is a well-known company.
	op2	My relatives and people close or important to me believe that [Company X] is a highly respected company.
	op3	My relatives and people close or important to me believe that [Company X] is an admired company.
	op4	My relatives and people close or important to me believe that [Company X] is a prestigious company.
	op5	People in general think that [Company X] is a well-known company.
	op6	People in general think that [Company X] is a highly respected company.
	op7	People in general think that [Company X] is an admired company.
	op8	People in general think that [Company X] is a prestigious company.
OI	oi1	When someone criticizes [Company X] it feels like a personal insult.
	oi2	I am very interested in what others think about [Company X].
	oi3	When I talk about [Company X], I usually say "we" rather than "they".
	oi4	[Company X's] successes are my successes.
	oi5	When someone praises [Company X] it feels like a personal compliment.
	oi6	If a story in the media criticized [Company X], I would feel embarrassed.
AC_J	acj1	I would be very happy to spend the rest of my career with [Company X].
	acj2	I enjoy discussing [Company X] with people outside of it.
	acj3	I really feel the problems of [Company X] are my own.
	acj4	[Company X] has a great deal of personal meaning for me.
AC_L	acl1	I do not feel like part of a family at [Company X].
	acl2	I do not feel emotionally attached to [Company X].
	acl3	I do not feel a strong sense of belonging to [Company X].

¹ Note that OP = organizational prestige, OI = organizational identification, AC_J = affective commitment (joy), and AC_L = affective commitment (love).

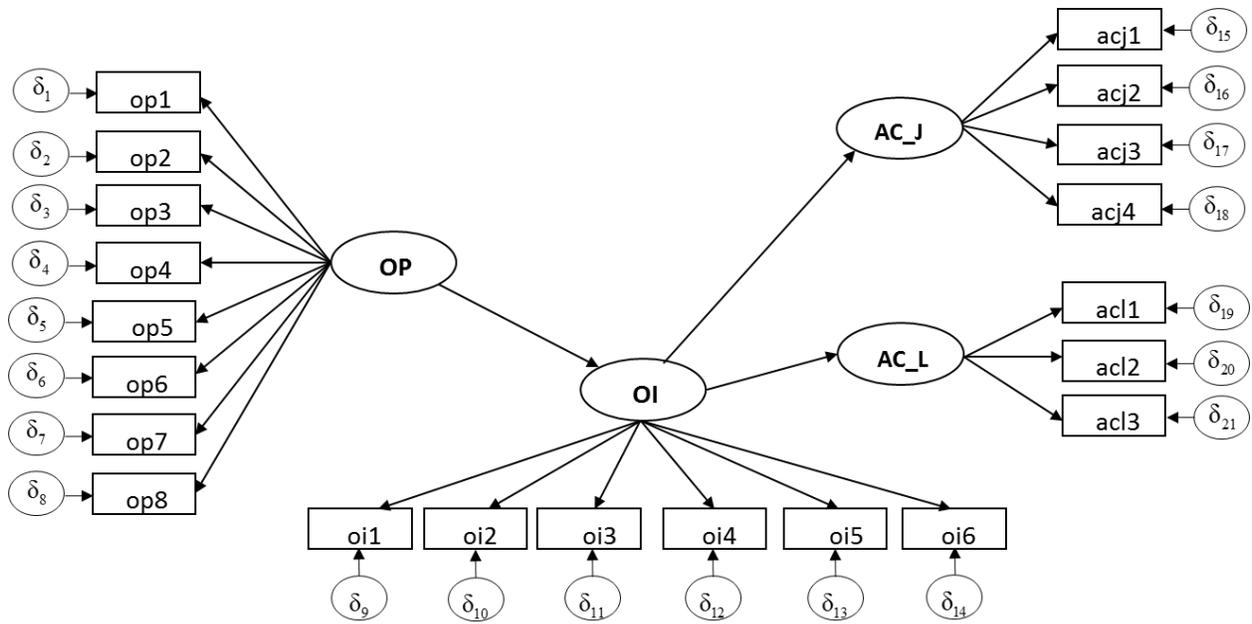


Figure 6- 1. A hypothesized model for the organizational identification data

6.2. Analysis

BGSCA was applied to fit the specified model to the example data. As discussed in Chapter 4, the Gibbs sampler algorithm was used to obtain the posterior distributions for the model parameters. Given little prior knowledge on the parameters, we used conjugate prior distributions with moderate-to-large variance hyperparameters. Specifically, for $\bar{\mathbf{z}}_i$, $\bar{\mathbf{W}}$, \mathbf{C} , and \mathbf{V} , normal distributions were assumed with mean hyperparameter value of zero and variance hyperparameter of ten. Prior distributions for $\phi_1^{(x)}, \dots, \phi_{J_1}^{(x)}, \phi_1^{(y)}, \dots, \phi_{J_2}^{(y)}$, and $\phi_1^{(\gamma)}, \dots, \phi_{p_2}^{(\gamma)}$ were assumed to follow inverse-gamma distributions with a relatively large variance as follows. Given

that the mean and variance of an inverse-gamma distribution were computed as $\frac{\nu}{a-1}$ and

$\frac{\nu^2}{(a-1)^2(a-2)}$ (where a is the shape and ν is the scale parameter of the distribution), respectively;

we set the shape parameter as five and the scale parameter as eight. In the analysis, the first 1000

iterations were discarded as the burn-in sample size, after which another set of 6000 iterations were run while saving samples at every second iteration from the algorithm for the posterior inferences. For comparison, the original GSCA was also applied to the same data using the R package `gesca`.

6.3. Results

Figures 6- 2 and 6- 3 display the trace and ACF plots of the estimates of several parameters only because those of the remaining parameter estimates showed similar patterns. A trace plot is to visualize how samples drawn from a posterior distribution for a parameter change over iterations. As shown in Figure 6- 2, the estimates of each parameter sampled over iterations did not show a systematic trend fluctuating around the center of its MCMC chain. An ACF plot can be used for evaluating whether or not the correlation between adjacent samples decays toward zero as the time lag increases. As shown in Figure 6- 3, the autocorrelation values for all the parameter estimates were very close to zero at either the first or second lag. Thus, overall, the Gibbs Sampler algorithm used in the analysis reached convergence.

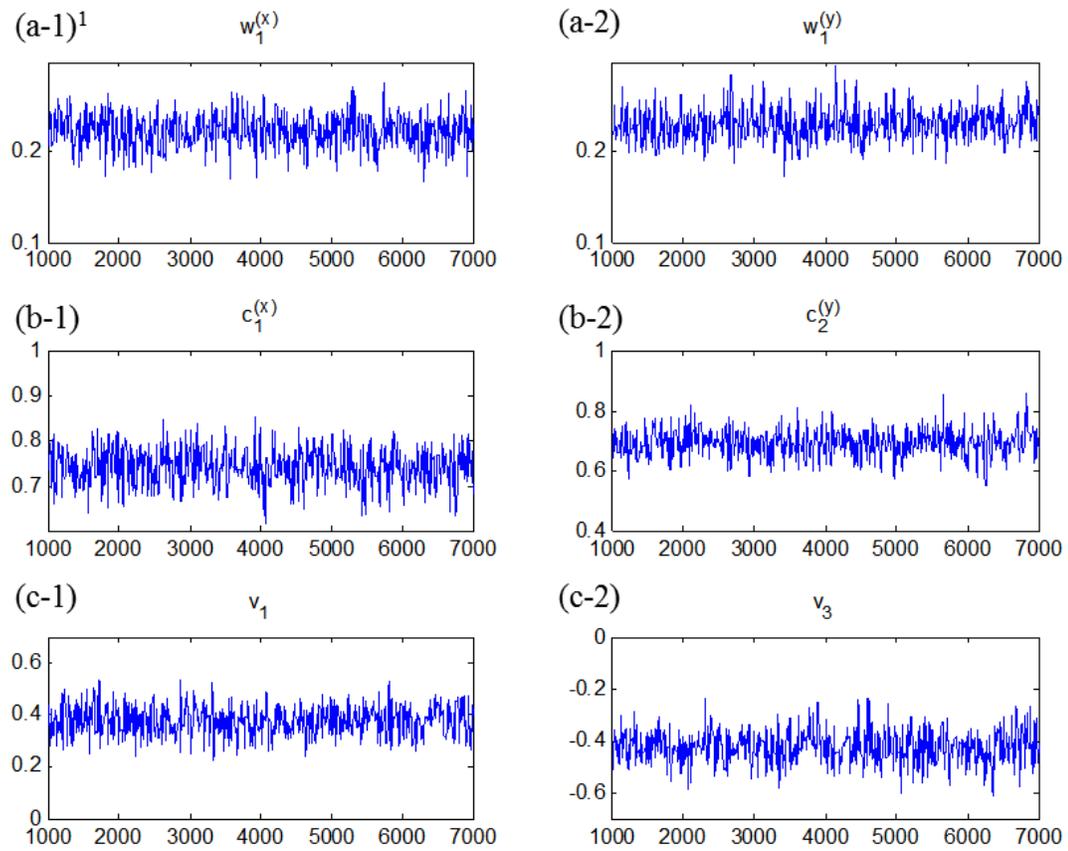


Figure 6- 2. Trace plots of the estimates of (a) weights, (b) loadings, and (c) path coefficients.
¹ (a-1) is a trace plot of $w_1^{(x)}$ (op1 \rightarrow OP); (a-2) $w_1^{(y)}$ (oi1 \rightarrow OI); (b-1) $c_1^{(x)}$ (OP \rightarrow op1); (b-2) $c_2^{(y)}$ (OI \rightarrow oi2); (c-1) v_1 (OP \rightarrow OI); (c-3) v_3 (OI \rightarrow AC_L).

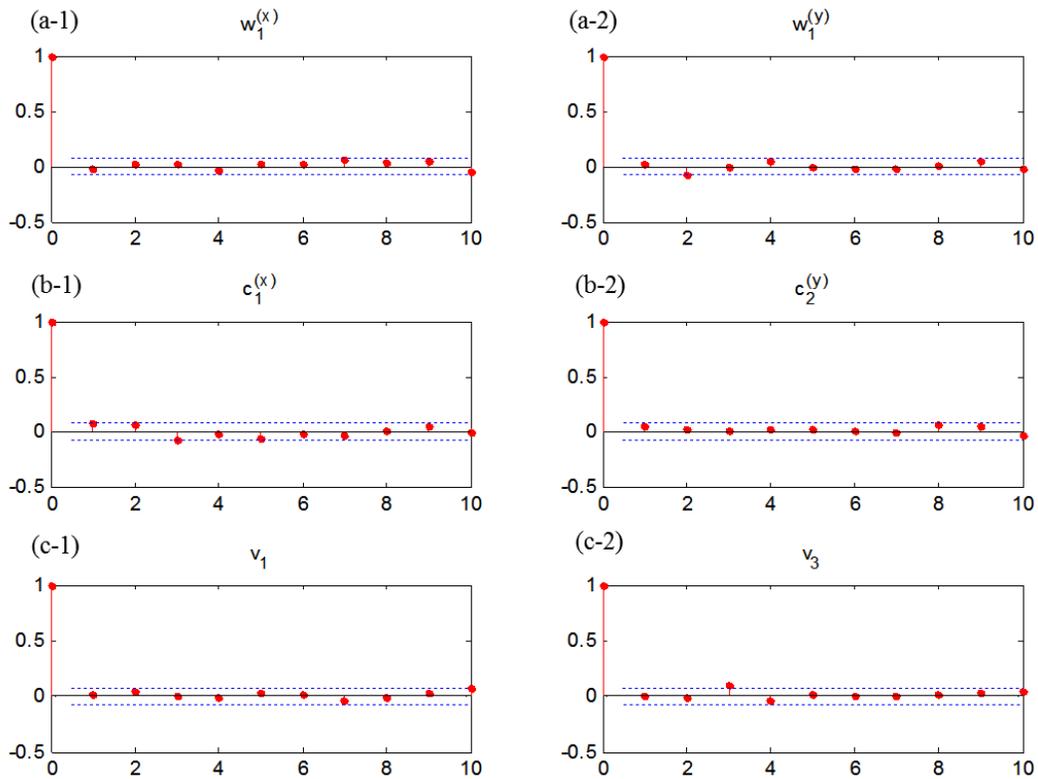


Figure 6- 3. ACF plots of the estimates of (a) weights, (b) loadings, and (c) path coefficients. (a-1) is an ACF plot of $w_1^{(x)}$ (op1 \rightarrow OP); (a-2) $w_1^{(y)}$ (oi1 \rightarrow OI); (b-1) $c_1^{(x)}$ (OP \rightarrow op1); (b-2) $c_2^{(y)}$ (OI \rightarrow oi2); (c-1) v_1 (OP \rightarrow OI); (c-2) v_3 (OI \rightarrow AC_L).

BGSCA resulted in $PPP_b = 0.60$, indicating that the chi-square statistic value of the observed data was less likely to be significantly different from that of replicated data drawn based on the resultant posterior estimates, because $PPP_b = 0.50$ is considered a good fit, as discussed in Chapter 3. Accordingly, the hypothesized model seemed to fit acceptably well to the data. The original GSCA estimation provided $GFI = 0.99$ and $SRMR = 0.08$. Given that a GFI value close to one and an SRMR value close to zero are considered a good fit (Hwang & Takane, 2014, Chapter 2), both GSCA and BGSCA suggested a good model fit. To describe the resultant posterior distributions obtained from BGSCA, the means, standard deviations (SD), medians of

the posterior distributions were calculated. Moreover, their credible intervals (C.I.) were computed to provide regions in which 95% of the posterior distributions for the parameters lie.

Table 6- 2 summarizes the results of the posterior distributions for the weight parameters. In the table, the last column contains the weight estimates obtained from original GSCA. The weight estimates show the contribution of each variable to defining its latent variable. As shown in Table 6- 2, the weight estimates for all observed variables per latent variable were similar to one another and statistically significant, indicating that the observed variables contributed equally well to determining their latent variables. Both BGSCA and original GSCA led to the same interpretations, although BGSCA tended to produce larger weight estimates than those from original GSCA.

Table 6- 2. Results of the weight estimates obtained from BGSCA.

Parameter	BGSCA				Original GSCA	
	Mean	SD	Median	C.I.		
Weights	op1 → OP	0.2225	0.018	0.2221	[0.1843 0.2589]	0.1502
	op2 → OP	0.2367	0.016	0.2365	[0.2034 0.2711]	0.1597
	op3 → OP	0.2179	0.018	0.2183	[0.1814 0.2551]	0.1570
	op4 → OP	0.2287	0.017	0.2281	[0.1949 0.2643]	0.1470
	op5 → OP	0.2304	0.017	0.2304	[0.1974 0.2713]	0.1619
	op6 → OP	0.2454	0.016	0.2458	[0.2125 0.2778]	0.1683
	op7 → OP	0.2210	0.018	0.2211	[0.1837 0.2586]	0.1503
	op8 → OP	0.2268	0.017	0.2265	[0.1895 0.2620]	0.1544
oi1 → OI	0.2965	0.024	0.2963	[0.2478 0.3450]	0.2191	
oi2 → OI	0.2782	0.026	0.2788	[0.2277 0.3274]	0.2109	
oi3 → OI	0.2192	0.003	0.2193	[0.1605 0.2774]	0.1944	
oi4 → OI	0.3224	0.023	0.3227	[0.2758 0.3689]	0.2610	
oi5 → OI	0.3208	0.023	0.3211	[0.2760 0.3652]	0.2373	
oi6 → OI	0.2753	0.026	0.2746	[0.2242 0.3244]	0.1840	
acj1 → AC_J	0.3700	0.039	0.3706	[0.2893 0.4443]	0.3024	
acj2 → AC_J	0.4060	0.037	0.4057	[0.3338 0.4802]	0.3296	
acj3 → AC_J	0.4313	0.037	0.4326	[0.3577 0.5058]	0.3645	
acj4 → AC_J	0.3343	0.042	0.3356	[0.2454 0.4197]	0.3035	
acl1 → AC_L	0.5136	0.061	0.5149	[0.3888 0.6412]	0.4531	
acl2 → AC_L	0.4461	0.056	0.4474	[0.3243 0.5670]	0.3874	
acls3 → AC_L	0.4861	0.060	0.4874	[0.3643 0.6048]	0.4663	

Table 6- 3. Results of the loading estimates obtained from BGSCA.

Parameter	BGSCA				Original GSCA	
	Mean	SD	Median	C.I.		
Loadings	OP → op1	0.7413	0.041	0.6589	[0.7403 0.8219]	0.7806
	OP → op2	0.7883	0.037	0.7129	[0.7903 0.8599]	0.8247
	OP → op3	0.7267	0.042	0.6797	[0.7252 0.8432]	0.7699
	OP → op4	0.7687	0.038	0.6844	[0.7687 0.8432]	0.8037
	OP → op5	0.7673	0.041	0.6841	[0.7662 0.8482]	0.8014
	OP → op6	0.8112	0.036	0.7416	[0.8113 0.8819]	0.8430
	OP → op7	0.7344	0.043	0.6494	[0.7361 0.8150]	0.7764
	OP → op8	0.7606	0.039	0.6844	[0.7592 0.8398]	0.8010
OI → oi1	0.7284	0.042	0.6443	[0.7276 0.8073]	0.7870	
OI → oi2	0.6904	0.045	0.5997	[0.6908 0.7850]	0.7580	
OI → oi3	0.5392	0.034	0.4315	[0.5396 0.6458]	0.6366	
OI → oi4	0.7973	0.038	0.7201	[0.7383 0.8741]	0.8234	
OI → oi5	0.7937	0.039	0.7102	[0.7939 0.8741]	0.8107	
OI → oi6	0.6821	0.047	0.5955	[0.6821 0.7794]	0.7430	
AC_J → acj1	0.6247	0.053	0.5169	[0.6254 0.7273]	0.7480	
AC_J → acj2	0.6889	0.051	0.5848	[0.6913 0.7886]	0.7900	
AC_J → acj3	0.7284	0.048	0.6333	[0.7291 0.8227]	0.8199	
AC_J → acj4	0.5719	0.049	0.4572	[0.5739 0.6782]	0.7072	
AC_L → acl1	0.6278	0.061	0.4960	[0.6268 0.6812]	0.7959	
AC_L → acl2	0.5445	0.056	0.4158	[0.5407 0.6805]	0.7094	
AC_L → acl3	0.5859	0.068	0.4567	[0.5861 0.7149]	0.7817	

The means, SD, medians, and 95% C.I. of the loading estimates are given in Table 6- 3.

All four latent variables were positively and statistically significantly related to their corresponding observed variables. OP was highly linked to all the eight observed variables having a mean loading estimate higher than 0.70, but in particular its association was most salient with op6 ($c_6^{(x)} = 0.81$, C.I. = [0.81 0.88]). OI had the highest association with the variables oi4 ($c_4^{(y)} = 0.80$, C.I. = [0.74 0.87]) and oi5 ($c_5^{(y)} = 0.79$, C.I. = [0.79 0.87]), followed by oi1 ($c_1^{(y)} = 0.73$, C.I. = [0.73 0.81]), oi2 ($c_2^{(y)} = 0.69$, C.I. = [0.69 0.79]), oi6 ($c_6^{(y)} = 0.68$, C.I. = [0.68 0.78]), and oi3 ($c_3^{(y)} = 0.54$, C.I. = [0.54 0.65]). The observed variable acj3 was more highly loaded on AC_J ($c_9^{(y)} = 0.73$, C.I. = [0.73 0.82]) than the remaining three observed

variables. Although all the loading estimates in related to AC_J were found to be statistically significant, the magnitude of the effects on acj1 ($c_7^{(y)} = 0.62$, C.I. = [0.63 0.73]) and acj4 ($c_{10}^{(y)} = 0.57$, C.I. = [0.57 0.68]) were relatively weak. AC_L was most highly related to acl1 ($c_{11}^{(y)} = 0.73$, C.I. = [0.63 0.68]), followed by acl3 ($c_{13}^{(y)} = 0.59$, C.I. = [0.59 0.71]) and acl2 ($c_{12}^{(y)} = 0.63$, C.I. = [0.63 0.68]). Compared to original GSCA, BGSCA yielded smaller loading estimates. This may be due to the fact that BGSCA obtained the estimates under the explicit specification of measurement error variances, as in factor-based SEM (see Chapter 4). It is well known that factor-based analyses involving such model specifications tend to provide smaller loading estimates than component-based analyses (e.g., Chin, 1995; Snook & Gorsuch, 1989; Widaman, 2007).

Table 6- 4 presents the means, standard deviations, medians, and 95% C.I. of the posterior distributions for the path coefficients. As expected, OP had a positive and statistically significant effect on OI ($v_1 = 0.38$, C.I. = [0.38 0.49]). This implied that employees with higher levels of OP were more likely to identify themselves with their organization. The effect of OI on AC_J was found to be statistically significant and positive ($v_2 = 0.61$, C.I. = [0.61 0.72]), suggesting that when employees were more likely to identify themselves with the organization, they were more intrinsically motivated and likely to internalize values of the organization. The effect of OI on AC_L was statistically significant and negative ($v_3 = -0.42$, C.I. = [-0.42 -0.30]), indicating that as employees were more inclined to identify themselves with the organization, they tended to have stronger psychological attachment to the workplace. Compared to original GSCA, BGSCA resulted in slightly larger path coefficients estimates. Again, factor-based analyses involving the specification of measurement error variances are expected to yield larger

path coefficient estimates than component-based analyses (e.g., Chin, 1995; Dijkstra & Henseler, 2015; Lu, Kwan, Thomas, & Cedzynski, 2011).

Table 6- 4. Results of the path coefficient estimates obtained from BGSCA.

Parameter	BGSCA				Original GSCA	
	Mean	SD	Median	C.I.		
Path	OP → OI	0.3766	0.057	0.2680	[0.3777 0.4852]	0.3615
coefficients	OI → AC_J	0.6144	0.053	0.5069	[0.6140 0.7164]	0.6138
	OI → AC_L	-0.4225	0.061	-0.5411	[-0.4241 -0.2966]	-0.4041

As discussed in Chapter 3, BGSCA allows comparing two (or more) competing models, using the Bayes factor (K_{01}). For an illustrative purpose only, suppose that the first model M_0 is the hypothesized model in Figure 6- 1 and the second model M_1 is the one displayed in Figure 6- 4. M_1 added the direct effects of *organizational prestige* on both *affective commitment_joy* and *affective commitment_love* to M_0 .

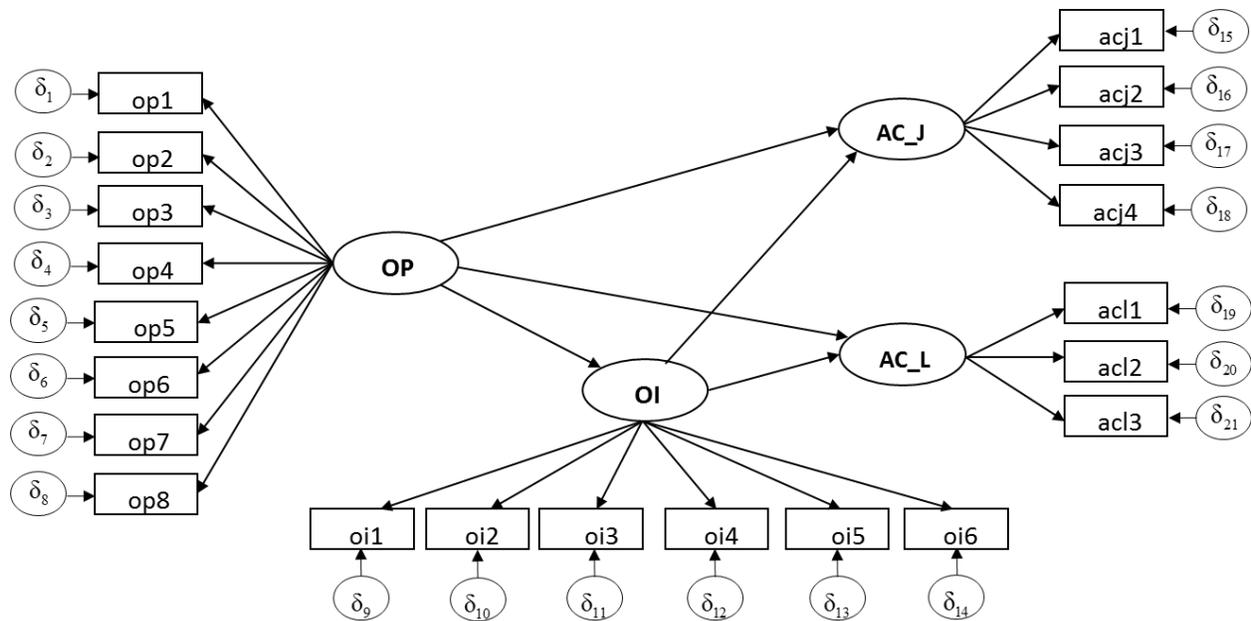


Figure 6- 4. A hypothetical model specified for model comparison.

The Bayes factor K_{01} measures the degree to which a model predicts the data better than the other competing model, Table 6- 5 provides the interpretation of different values of K_{01} , based on Kass and Raftery (1995). In this example, BGSCA provided $K_{01} = 7.3891$ (or equivalently $2\ln(K_{01}) = 4$), indicating that the data favored M_0 by a factor of about 7.39. Thus, we might conclude that M_0 was more supported by the data than M_1 . In the original GSCA, model comparison was carried out by comparing the AFIT values of the two competing models: the larger difference in the AFIT values indicates stronger evidence for a model with the higher AFIT value. The original GSCA provided the AFIT of $M_0 = 0.5322$ and that of $M_1 = 0.5335$. Although M_1 provided a higher AFIT value, the difference between the two values was trivial ($M_1 - M_0 = 0.0013$), and so we might also choose a more parsimonious model M_0 as the final one. Although both BGSCA and original GSCA reached the same conclusion, BGSCA enabled to provide an explicit measure that quantifies the relative strength in supporting a model against another.

Table 6- 5. Different Bayes factors and their interpretations.

K_{01}	$2\ln(K_{01})$	Interpretation (compared to M_1)
1 – 3	0 – 2	No or week evidence for
3 – 20	2 – 6	Positive evidence for M_0
20 – 150	6 – 10	Strong evidence for M_0
> 150	> 10	Very strong evidence for M_0

Chapter 7. Summary and Discussion

Bayesian GSCA or BGSCA was proposed to integrate the original GSCA into a Bayesian framework. As in the original GSCA, BGSCA combines three sub-models - measurement, structural, and weighted relation models, in a unified manner. Unlike the original GSCA, however, BGSCA estimates the true measurements of observed data, called error-free data, in order to discard additive noises or random errors given in the observed data and construct the weighted composites of the error-free data as proxies for latent variables. BGSCA also permits the incorporation of any prior knowledge on a parameter, in the form of a prior distribution, into the modeling process before observing the data, which is in turn combined with the likelihood function of the data. In BGSCA, Gibbs Sampler is used to update the posterior distributions for the parameters. Accordingly, the samples generated from each posterior distribution are used to calculate the mean, median, standard deviation, and 95% credible interval of the posterior distribution.

The performance of BGSCA was examined based on both simulated and real data sets. The first simulation study discussed in Chapter 5 showed that BGSCA recovered parameters reasonably well under various conditions. The second simulation study was to examine the sensitivity or robustness of BGSCA results under different specifications of prior distributions. BGSCA was also applied to fit a hypothesized model to a real data set. BGSCA generally yielded solutions similar to those obtained from the original GSCA, leading to the same interpretations. Nonetheless, it was able to explicitly take into account measurement errors of observed variables, which were neglected in the original GSCA, so that it seemed to circumvent the propensity of overestimating loadings and underestimating path coefficients in the original GSCA when the errors are present.

Despite the technical and empirical implications, BGSCA can be further extended to improve its scope of applicability. The current BGSCA employs the standard GSCA model that assumes only linear relations among latent variables. However, researchers may be interested in examining nonlinear relationships among latent variables. In the literature, such nonlinear SEM has received considerable attention (e.g., Henseler, & Chin, 2010; Kenny & Judd, 1984; Schumacker, & Marcoulides, 1998). For example, when an interaction term between the latent variables $\gamma_{i1}^{(x)}$ and $\gamma_{i2}^{(x)}$ in Figure 2- 1 is assumed, the structural model would be specified as

$$\gamma_{i1}^{(y)} = b_1 \gamma_{i1}^{(x)} + b_2 \gamma_{i2}^{(x)} + b_3 \gamma_{i1}^{(x)} \gamma_{i2}^{(x)} + \zeta_i. \quad (7.1)$$

To accommodate various forms of nonlinearities in general, including both interaction and quadratic terms of latent variables, (7.1) can be rewritten as follows:

$$\gamma_{i1}^{(y)} = b_1 \gamma_{i1}^{(x)} + b_2 \gamma_{i2}^{(x)} + \mathbf{b}_{(3)}^* \mathbf{F}(\gamma_{i1}^{(x)}, \gamma_{i2}^{(x)}) + \zeta_i, \quad (7.2)$$

where $\mathbf{b}_{(3)}^*$ is a vector of path coefficients of the nonlinear latent variables affecting the endogenous latent variable, and $\mathbf{F}(\gamma_{i1}^{(x)}, \gamma_{i2}^{(x)})$ is a vector of nonlinear functions of $\gamma_{i1}^{(x)}$ and $\gamma_{i2}^{(x)}$.

For example, for the model in (7.1), the third term in (7.2) is expressed as

$$\mathbf{b}_{(3)}^* \mathbf{F}(\gamma_{i1}^{(x)}, \gamma_{i2}^{(x)}) = \begin{bmatrix} b_{(3),1}^* & b_{(3),2}^* & b_{(3),3}^* \end{bmatrix} \begin{bmatrix} \gamma_{i1}^{(x)} \gamma_{i2}^{(x)} \\ \gamma_{i1}^{(x)2} \\ \gamma_{i2}^{(x)2} \end{bmatrix} \quad (7.3)$$

A promising way of extending BGSCA to handle these nonlinear relations is to adopt a similar method proposed by Guo, Zhu, Chow, and Ibrahim (2012). They approximate nonlinear functions of latent variables by a linear combination of known basis functions, which is called the basis function expansion approach. Specifically, each nonlinear term in $\mathbf{F}(\gamma_{i1}^{(x)}, \gamma_{i2}^{(x)})$ in (7.2) can be approximated by a weighted sum of tensor product basis functions as follows:

$$\mathbf{f}(\gamma_{i1}^{(x)}, \gamma_{i2}^{(x)}) = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \beta_{m_1 m_2} \psi_{1m_1}(\gamma_{i1}^{(x)}) \psi_{2m_2}(\gamma_{i2}^{(x)}), \quad (7.4)$$

where ψ_{1m_1} and ψ_{2m_2} refer to the basis functions for the first and second exogenous latent variable, respectively, and $\beta_{m_1 m_2}$ indicates the basis coefficient of the m_1 and m_2 th basis functions.

Moreover, a Bayesian Lasso method (Park & Casella, 2008) is used to address potential overfitting and accomplish variable selection simultaneously, extracting only a subset of exogenous latent variables whose nonlinear terms would have strong associations with endogenous latent variables. This method can readily be combined with BGSCA, enabling to capture various forms of nonlinearities more effectively and at the same time, to select a subset of exogenous latent variables.

The abovementioned basis function expansion can be applied even further for the analysis of functional data that are collected in the form of smooth curves or functions varying over a continuum such as time or space (Ramsay & Silverman, 2005). To accommodate functional data, the observed variables and loadings in the measurement model of BGSCA would be replaced by data functions and loading functions, respectively. There are generally two Bayesian approaches to the analysis of functional data: One has been built upon the conventional basis function expansion but extended such that selecting the number and/or locations of the knots for the basis functions (i.e., a sequence of breakpoint values that divide the entire interval into subintervals) are considered the parameters to be estimated (DiMatteo, Genovese, & Kass, 2001; Kohn, Smith, & Chan, 2001; Lindstorm, 1999). The other approach is to obtain smooth functions using a nonparametric Bayesian model, in which parameters are assumed to be spanned on an infinite dimensional space, called stochastic processes (e.g., Behseta, Kass, & Wallstrom, 2005; Ferraty & Vieu, 2006; Rodríguez, Dunson, & Gelfand, 2009). Although both

approaches seem to be applied to BGSCA, future studies are warranted to investigate the technical and empirical feasibility of these extensions.

In the current BGSCA, parameters are estimated under the assumption that all observations in data are collected from a single population. This is considered *aggregate sample analyses* in that parameters are estimated by pooling the data across all observations. Nonetheless, such homogeneity assumption is often challenging to satisfy, and instead it may be more realistic to assume that observations are compounded from a number of heterogeneous subgroups, each of which shows distinct relationships between independent and dependent variables. The importance of distinguishing such heterogeneous subgroups has been emphasized in various studies, for instance, in modeling brand choice decision (e.g., Kamakura & Russell, 1989; Kamakura, Kim, & Lee, 1996), customer satisfaction (e.g., Swait & Sweeney, 2000), and cognitive strategies in children's executive control processes (e.g., Dauvier, Chevalier, & Blaye, 2012). To partition observations into heterogeneous subgroups, it would be worthwhile to consider combining BGSCA with Bayesian mixture models (e.g., Dieblot & Robert, 1994; Neal, 1992; Richardson & Green, 1997; Svensén & Bishop, 2005). This mixture extension can be regarded as an alternative to fuzzy clusterwise GSCA (Hwang et al., 2007), because it would allow selecting the optimal number of subgroups automatically. With fuzzy clusterwise GSCA, the number of subgroups needs to be empirically selected by examining how the global optimization criterion and/or other cluster validity measures change against different numbers of subgroups, whereas with the Bayesian mixture model, the number of subgroups can be treated as another parameter and hence estimated via an MCMC method.

References

- Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of occupational and organizational psychology*, 63(1), 1-18.
- Ansari, A., Jedidi, K., & Jagpal, S. (2000). A hierarchical Bayesian methodology for treating heterogeneity in structural equation models. *Marketing Science*, 19(4), 328-347.
- Arminger, G., & Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, 63(3), 271-300.
- Behseta, S., Kass, R. E., & Wallstrom, G. L. (2005). Hierarchical models for assessing variability among functions. *Biometrika*, 92(2), 419-434.
- Bergami, M., & Bagozzi, R. P. (2000). Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization. *British Journal of Social Psychology*, 39(4), 555-577.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.
- Berger, J. O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of statistical planning and inference*, 25(3), 303-328.
- Bishop, C. M. (1999). Bayesian PCA. *Advances in neural information processing systems*, 11, 382-388.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: A least-squares estimator. *Sociological methodology*, 25, 223-252.

-
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* (Vol. 467): John Wiley & Sons.
- Bolstad, W. M. (2007). *Introduction to Bayesian statistics*. John Wiley & Sons.
- Bouriga, M., & Féron, O. (2013). Estimation of covariance matrices based on hierarchical inverse-Wishart priors. *Journal of Statistical Planning and Inference*, 143(4), 795-808.
- Box, G. E. P., & Tiao, G. C. (1992). *Bayesian inference in statistical analysis*. New York: Wiley.
- Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall/CRC.
- Carlin, B. P., & Louis, T. A. (2008). *Bayesian methods for data analysis*. CRC Press.
- Chen, H., Bakshi, B. R., & Goel, P. K. (2009). Integrated estimation of measurement error with empirical process modeling—A hierarchical Bayes approach. *AIChE journal*, 55(11), 2883-2895.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90(432), 1313-1321.
- Chib, S., & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4), 327-335.
- Chin, W. W. (1995). Partial least squares is to LISREL as principal components analysis is to common factor analysis. *Technology Studies*, 2(2), 315-319.
- Chin, W. W. (2001). *PLS-Graph user's guide, Version 3.0*. Houston, TX: Soft Modelling.
- Chou, C. P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: a Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, 44(2), 347-357.
- Congdon, P. (2007). *Bayesian statistical modelling* (Vol. 704): John Wiley & Sons.

- Dauvier, B., Chevalier, N., & Blaye, A. (2012). Using finite mixture of GLMs to explore variability in children's flexibility in a task-switching paradigm. *Cognitive Development*, 27(4), 440-454.
- de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41(4), 471-503.
- Depaoli, S. (2014). The impact of inaccurate “informative” priors for growth parameters in Bayesian growth mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 239-252.
- Diebolt, J., & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(2), 363-375.
- Dijkstra, T. K., & Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational statistics & data analysis*, 81, 10-23.
- DiMatteo, I., Genovese, C. R., & Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, 88(4), 1055-1071.
- Edwards, W., Lindman, H., & Savage, L. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70(3), 193.
- Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*: Springer Science & Business Media.
- Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Muller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, Connecticut: IAP.

-
- Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3), 501-514.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (Vol. 2). Boca Raton, FL: Chapman & Hall/CRC.
- Gelman, A., & Meng, X. L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, 13(2), 163-185.
- Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, 6(4), 733-760.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6, 721-741.
- Ghosh, J. K., Delampady, M., & Samanta, T. (2007). *An introduction to Bayesian analysis: theory and methods*: Springer Science & Business Media.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice*. London: Chapman & Hall.
- Guo, R., Zhu, H. T., Chow, S. M., & Ibrahim, J. G. (2012). Bayesian lasso for semiparametric structural equation models. *Biometrics*, 68(2), 567-577.
- Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *The Annals of Statistics*, 8(3), 586-597.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97-109.

- Hau, K. T., & Marsh, H. W. (2004). The use of item parcels in structural equation modelling: Non-normal data and small sample sizes. *British Journal of Mathematical and Statistical Psychology*, *57*(2), 327-351.
- Henseler, J., & Chin, W. W. (2010). A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. *Structural Equation Modeling*, *17*(1), 82-109.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*: Springer Science & Business Media.
- Hwang, H. (2009). Regularized generalized structured component analysis. *Psychometrika*, *74*, 514-530.
- Hwang, H., Desarbo, W. S., & Takane, Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika*, *72*(2), 181-198.
- Hwang, H., Ho, R. M., & Lee, J. (2010). Generalized structured component analysis with latent interactions. *Psychometrika*, *75*(2), 228-242.
- Hwang, H., Kim, S., Lee, S., & Park, T. (2016). gesca: Generalized Structured Component Analysis (GSCA). R package version 1.0.1. <https://CRAN.R-project.org/package=gesca>.
- Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, *69*(1), 81-99.
- Hwang, H., & Takane, Y. (2010). Nonlinear Generalized Structured Component Analysis. *Behaviormetrika*, *37*(1), 1-14.
- Hwang, H., & Takane, Y. (2014). *Generalized structured component analysis: A component-based approach to structural equation modeling*: CRC Press.

-
- Hwang, H., Takane, Y., & Malhotra, N. (2007). Multilevel generalized structured component analysis. *Behaviormetrika*, *34*(2), 95-109.
- Hwang, H., Takane, Y., & Tenenhaus, A. (2015). An alternative estimation procedure for partial least squares path modeling. *Behaviormetrika*, *42*(1), 63-78.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183-202
- Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57-88). Mahwah, NJ: Erlbaum.
- Jung, K., Takane, Y., Hwang, H., & Woodward, T. S. (2012). Dynamic GSCA (Generalized Structured Component Analysis) with applications to the analysis of effective connectivity in functional neuroimaging data. *Psychometrika*, *77*(4), 827-848.
- Kamakura, W. A., Kim, B. D., & Lee, J. (1996). Modeling preference and structural heterogeneity in consumer choice. *Marketing Science*, *15*(2), 152-172.
- Kamakura, W. A., & Russell, G. (1989). A probabilistic choice model for market segmentation and elasticity structure.
- Kaplan, D., & Depaoli, S. (2012). Bayesian statistical methods. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650-673). New York: Guilford Press.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

- Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90(431), 928-934.
- Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, 96(1), 201.
- Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65(4), 457-474.
- Kohn, R., Smith, M., & Chan, D. (2001). Nonparametric regression using linear combinations of basis functions. *Statistics and Computing*, 11(4), 313-322.
- Lee, P. M. (2012). *Bayesian statistics: an introduction*. Chichester, West Sussex: Wiley.
- Lee, S. Y. (2006). Bayesian analysis of nonlinear structural equation models with nonignorable missing data. *Psychometrika*, 71(3), 541-564.
- Lee, S. Y. (2007). *Structural equation modeling a Bayesian approach*. New York: Wiley.
- Lee, S. Y., & Song, X. Y. (2004). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, 57(1), 131-150.
- Lee, S. Y., & Song, X. Y. (2012). *Basic and advanced structural equation modeling with applications in the medical and behavioral sciences*. Hoboken: Wiley.
- Lee, S. Y., & Zhu, H. T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, 53(2), 209-232.

-
- Lewis, S. M., & Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace—Metropolis estimator. *Journal of the American Statistical Association*, 92(438), 648-655.
- Lindstrom, M. J. (1999). Penalized estimation of free-knot splines. *Journal of Computational and Graphical Statistics*, 8(2), 333-352.
- Lohmöller, J. B. (1984). *LVPLS Program manual Version 1.6: Latent Variables path analysis with partial least-squares estimation*. Zentralarchiv für Empirische Sozialforschung. Universitst zu Koln.
- Lohmöller, J. B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg: Physica.
- Lu, I. R. R., Kwan, E., Thomas, D. R., & Cedzynski, M. (2011). Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing*, 28(3), 258-268.
- Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York: Springer
- Mael, F., & Ashforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of organizational Behavior*, 13(2), 103-123.
- McDonald, R. P. (1996). Path analysis with composite variables. *Multivariate Behavioral Research*, 31(2), 239-270.
- Meng, X. L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 6, 831-860.

- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6), 1087-1092.
- Moustaki, I. (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology*, 56(2), 337-357.
- Mulaik, S. A. (2009). *Linear causal modeling with structural equations*: CRC Press.
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological methods*, 17(3), 313.
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189.
- Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115-132.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological methods & research*, 22(3), 376-398.
- Neal, R. M. (1992). Bayesian mixture modeling. In C. R. Smith, G. J. Erickson, & P. O. Neudorfer (Eds.), *Maximum Entropy and Bayesian Methods: Proceedings of the 11th International Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis* (pp. 197-211). Dordrecht: Kluwer Academic Publishers.
- Newton, M. A., & Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(1), 3-48.

-
- Nounou, M. N., Bakshi, B. R., Goel, P. K., & Shen, X. (2002). Bayesian principal component analysis. *Journal of chemometrics*, *16*(11), 576-595.
- Palomo, J., Dunson, D. B., & Bollen, K. (2007). Bayesian structural equation modeling. In S. Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 163-188). Amsterdam: Elsevier.
- Park, T., & Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681-686.
- Press, S. J. (2003). *Subjective and objective Bayesian statistics: principles, models, and applications*: John Wiley & Sons.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis* (2nd ed.). New York: Springer.
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods*, *17*(3), 354.
- Richardson, S., & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)*, *59*(4), 731-792.
- Ringle, C.M., Wende, S., & Will, A., 2005. SmartPLS 2.0. Hamburg. www.smartpls.de.
- Robert, C. P. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*: Springer Science & Business Media.
- Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods*. New York: Springer.
- Rodríguez, A., Dunson, D. B., & Gelfand, A. E. (2009). Bayesian nonparametric functional data analysis through density estimation. *Biometrika*, *96*(1), 149-162.

- Satorra, A., & Muthen, B. O. (1995). Complex sample data in structural equation modeling. *Sociological methodology, 25*, 267-316.
- Scheines, R., Hoiijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika, 64*(1), 37-52.
- Schlaifer, R., & Raiffa, H. (1961). *Applied statistical decision theory*. Boston, MA: Division of Research, Graduate School of Business Administration, Harvard University.
- Schumacker, R. E., & Marcoulides, G. A. (1998). *Interaction and nonlinear effects in structural equation modeling*: Lawrence Erlbaum Associates Publishers.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461-464.
- Shimizu, S., & Kano, Y. (2008). Use of non-normality in structural equation modeling: Application to direction of causation. *Journal of Statistical Planning and Inference, 138*(11), 3483-3491.
- Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin, 106*(1), 148.
- Song, X. Y., & Lee, S. Y. (2004). Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology, 57*(1), 29-52.
- Song, X. Y., & Lee, S. Y. (2012). A tutorial on the Bayesian approach for analyzing structural equation models. *Journal of Mathematical Psychology, 56*(3), 135-148.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639.

-
- Spiegelhalter, D. J., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS user manual. Version 1.4. Cambridge: Medical Research Council Biostatistics Unit. Retrieved from <https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>
- Suk, H. W., & Hwang, H. (2016). Functional generalized structured component analysis. *Psychometrika*, *81*(4), 940-968.
- Svensén, M., & Bishop, C. M. (2005). Robust Bayesian mixture modelling. *Neurocomputing*, *64*, 235-252.
- Swait, J., & Sweeney, J. C. (2000). Perceived value and its impact on choice behavior in a retail setting. *Journal of Retailing and Consumer Services*, *7*(2), 77-88.
- Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational statistics & data analysis*, *48*(1), 159-205.
- Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(3), 611-622.
- Wentzell, P. D., Andrews, D. T., Hamilton, D. C., Faber, K., & Kowalski, B. R. (1997). Maximum likelihood principal component analysis. *Journal of Chemometrics*, *11*(4), 339-366.
- Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177-203). Mahwah, NJ: Lawrence Erlbaum.
- Wold, H. (1975). Path models with latent variables: The NIPALS approach. In H. M. Blalock, A.

- Aganbegian, F. M. Borodkin, R. Boudon, & V. Capecchi (Eds.), *Quantitative sociology: International perspectives on mathematical and statistical modeling* (pp. 307-357). New York: Academic.
- Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observations: Part II* (pp. 1-54). Amsterdam: North Holland.
- Wold, H. (1985). Partial least squares. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (pp. 581-591). New York: Wiley.