Research Journal of Mathematical and Statistical Sciences ISSN 2320–6047
Vol. **3(12),** 5-15, December **(2015)** *Res. J. Mathematical and Statistical Sci.*

Bayesian Analysis of Multiple Group Nonlinear Structural Equation Models with Ordered Categorical and Dichotomous Variables: A Survey

Thanoon Y. Thanoon[1,3] and Robiah Adnan[2]
[1]Department of Mathematical Science, Faculty of Science, University Teknologi Malaysia, 81310, Skudai, Johor, MALAYSIA
[3]Department of Operation Management Technique, Technical College of Management, Foundation of Technical Education, Mosul, IRAQ

Available online at: www.isca.in, www.isca.me
Received 8th October 2015, revised 20th October 2015, accepted 26th November 2015

Abstract

This paper is designed to give a complete overview of the literature that is available, as it relates to application of the Bayesian analysis model to investigate multiple group nonlinear structural equation models, also known as SEMs, including those having ordered categorical, dichotomous and categorical-dichotomous mixed variables. It will also work to summarize Bayesian multiple group nonlinear SEMs with nonlinear covariate variables, and latent variables in the structural model and both linear covariant and latent variable sin the measurement models. More specifically, it will be suggested that using hidden continuous normal distribution, including both right and left censoring and truncation, and interval censoring and truncation, can improve the Bayesian approach to multiple group nonlinear structural equation models when solving problems using ordered categorical and dichotomous data.

Keywords: Structural Equation Models, Bayesian analysis, Gibbs sampling, ordered categorical data, dichotomous data.

Introduction

Structural equation modeling is a statistical instrument commonly used in statistics in order to find the connection, or relationship, between the observed variable set and latent variable data. The observed variables are either manifest, also referred to "indicator variables", or they are latent variables, also known as "unobserved variables"[1].

The procedure for SEM is made up of two principle steps: causal process and structural relations. Casual procedures under study are made up of a series of structural equations. The structural relation is a pictorial representation that is used to create a clear definition of the hypothesis being studied. This allows the theoretical model to be examined, from a statistical vantage point, as part of the system of variables as a whole. This identifies the extent to which the data and the system are consistent. When goodness-of-it is deemed sufficient, then the model will assume the relation between variables,, otherwise the acceptability is rejected [2].

Using the SEM model poses multiple significant advantages. First, it allows the user to model more than one dependent variable at the same time. This saves significant time, and allows for more accurate comparison. It is also effective at testing the overall fit of the model, and is capable of managing both direct and indirect effects. It can also be used to model both complex hypotheses, specific hypotheses and parametric invariance in various subject-based groupings. Furthermore, SEM is such that it can be used for count, censored dichotomous, non-normal, and ordered categorical outcomes[3].

As such, there are multiple proposed methods for managing the problems that arise with ordered categorical and dichotomous data when placed in the traditional SEM model. Most commonly these include the weighted least squares model, the weight least squares model with both the mean and the variance adjusted, the robust maximum likelihood, the generalized least squares model, unweighted least squares, mean and variance adjusted maximum likelihood, and mean adjusted maximum likelihood. Furthermore, the asymptotically distribution free variable, also known as the weighted least squares chi-square test statistic, van be used in place of these options if all the outcome variables are continuous[4].

One methodology, proposed in a study by Lee et al.[5] to solve the problems that arise with ordered categorical data by using thresholds, or cut points. This means that by using the basis of observed ordinal categorical data, SEM, and CFA, the cut point is analyzed using the maximum likelihood approach. This method, more specifically, presents the idea that the NIL estimate of the parameters, latent variable estimates, and model comparison for the Bayesian information criterion and utilization of SEMs to quantify real data, like quality of life data.

Another, complimentary study by Song and Lee[6] presented the concept that a confirmary factor analysis model with covariates can be used to analyze dichotomous data when it is defined by a multivariate prohibit model. This is highly useful when analyzing data in the medical field. This provides a generalization of many beneficial multivariate prohibit models, and creates a very adaptable system for practical applications. As such, both the Monte-Carlo algorithm and the maximum likelihood estimation are both utilized to create a path sampling approach capable of calculating the observed data log, and also for evaluation the mode using the BIC for model comparison. Further, Fox[7] demonstrated that the SEM package in R program can be used to analyze the structural equation models. Further, this study showed that the integration of SEM packages and and other facilities into the R program is effective in fitting structural equations of the observed variable model via the two stage least square regression. It can also fit with the latent variable model by means of the full information maximum likelihood, which assumes multinormality.

Similarly, Asparouhov and Muth˙en[8] implemented use of the exploratory structural equation modeling (ESEM) by describing it as an exploratory factor analysis approach. In a similar set of findings, Yang and Green[9] worked to demonstrate the structural equation modeling (SEM) can be reliable when correctly managed, but that there are two major potential problems with the method employed: that the estimates may become unstable when the sample is too small, or when it has a bias created by mis-specified models. As a result a Monte Carl method was used to investigate the quality of the SEM estimates generated and the reliability of the coefficient alpha. Further, Iacobucci[10] worked to create some fit indices and developed advanced topics in SEM, in order to address certain advanced issues. This includes moderators, longitudinal data, mediation, fit incidences and sample size. This is complimented by Markus[11], which used casual explanations for SEM's in combination with structural equations to clarify these types of issues, and refine the use of SEM for casual explanation and thus clarified the behavioral science methodology behind those systems.

Hildreth[12] defined the significance of residual analysis in SEM. The finite sample and the asymototic properties of a class of residual estimators can be considered weighted functions of the observed variables, and can be derived from the SEM. Thus, the residual constructed using the proposed class of residual estimators can be analyzed in order to determine the outliers and the influential observations. Supportingly, Wu and Kwok [13] created a system of SEMs capable of analyzing complex survey data, and then comparing that data. The comparison is conducted concerning the design-based single level approach and model-based multilevel methodologies for analysis. Furthermore, Paul and Anderson[14] created a strategy for creating and testing casual models that depend on ordination axes taken from multivariate species data. These approaches differ from the previously described models, in that recent advancement in the casual modeling which have resulted in the ability to create and test SEMs free of limitation regarding both of functional forms and error distribution in the structural equation based model. Casual models of the effect were created and tried via the distance based redundancy analysis; it was also determined effective in forecasting the time it requires for the community to gain recovery, when a nonlinear model could be fixed to the PCO axes, or to fitted-nonlinear models.

Nonlinear Structural Equation Models

A non-linear SEM is created by implementing a measurement equation. This is essentially equivalent to using a linear SEM, but by using structural equations, which are nonlinear according to its exogenous latent variables. Thee conjectural reason for altering the model is naturally derived from the problem, because the SEM is closely related to the extension of the simple regression which has latent variables, as it relates to those that are multiple regressionary models with latent variables. The concept that nonlinear relations amid exogenous latent variables is key to the creation of more essential and more correct models in certain circumstances motives the further development of these nonlinear SEMs[15]. For example, Lee and Song [16]recommended a new nonlinear concept for the evaluation of SEMs with fixed covariates. A model designed to harness the power of path sampling for computing the Bayes factor was created in order to complete the model comparison.

To do so the required random observations were simulated according to the hybrid algorithm, using an amalgamation of the Gibbs sampling approach and the Metropolis- Hastings algorithm.

Lee and Tang[17] further presented his analysis of nonlinearSEMs, with covariates and both ordered and mixed continuous categorical results inspite of missing observations and absentee covariates, which are considered missing was a nonignorable mechanism. In this case, the nonignorable missingness mechanism is essential, and exclusive, to the logistic regression model. This is supported by the work of Lee et al.[18] which demonstrated that using the correlated continuous and discrete data sets in the nonlinear SEMs. Because correlated discrete data is commonly used in expedient applications, a nonlinear SEM that can be used with covariant as well as ordered categorical, unordered categorical, and mixed continuous variables is desired.

The Maximum likelihood method is used in order to generate the estimate and the model comparison estimation and model comparison. For example, the cardiovascular disease data set can be used to demonstrate this methodology. Not unsimilarly, Lee and Song[19] created a two-level, non-linear SEM model, which contained covariates. His method implemented maximum likelihood analysis, and suggested that a two-level approach, capable of managing the nonlinear, casual relationship between latent variables. This also provides insight into the effects of fixed covariates on these problem types.

A final development in the use of nonlinear SEMs was presented in a study by Henseler and Chin[20], who compared different methods for analyzing and interpreting the effects of different interactions between the latent variables including: the four partial least squares based approach, a product linear approach, a two staged approach, a hybrid approach and an orthogonalizing approach. These basic findings were complimented by the work of Wen et al.[21]which argued that the appropriate standardized solutions of SEMs for latent interactions which used these solutions to estimate the parameter of structural equation models were effective. Further, Codd [22] generated a body of applications of nonlinear structueal equation models with are relevant to psychological data, and used it to demonstrate how marginal maximum likelihood can be used to create an estimate of the general nonlinear structural equation models. Similarly, Pek et al.[23] created, and later evaluated, a delta method which focused on a parametric bootstrap approach for attaining estimated confidence intervals for Bauer's semiparametric methods for demonstrating nonlinear relations among latent variables.

Structural Equation Models with Multiple Group Data

Multiple Group Data, like that used with SEMs, is derived from a small number of groups with in the specified population. The total amount of observations in each grouping is generally both large and independent. As such, the main purpose of using the multiple group data analysis, as an approach, is to look into the similarities and differences between the models in the different groups. This results in a body of statistical inferences emphasized in the analysis of the multiple groups of the SEMS.

These are different from those that come from analyzing a two-level SEM. More specifically, the analysis of the multiple group SEM is a significant point of research because it can be used to investigate the behaviors of specific groups. For example this model could be used to determine the behaviors of teams of employees, different social groups, different treatment groups within a research study, and more. While the focal point of this study could be to test the hypotheses of diverse invariances among the models from different groups, the concept can also be effectively addressed as a model comparison problem, that is capable of addressing the Bayes factor or device information criteria for the Baysian approach. The benefit of selecting he Baysian approach over the Bayes factor or device information criteria is that the Baysian model comparison allows the researcher to compare non-nested models and there hypotheses. Thus, it is not always essential to follow the hierarchy of the hypotheses when assessing the invariance of the SEMs of different groups [15,3].

Bearing this in mind, a more current method was proposed by Lee[24] which uses the Bayesian analysis of nonlinear SEMs in order to solve the issues tat arise when using the nonignorable missing data and logistic regression. This is a hybrid algorithm that uses both the the Metropolis–Hastings algorithm and the Gibbs sampler in order to generate the joint Bayesian estimates for the latent variables, structural parameters, standard error estimates and parameters in the nonignorable missing model.

In a related study by Rabe-Hesketh et al.[25], it was determined that a unifying strategy for generalized, or nonspecific, multilevel SEMs could be derived from the models within the framework. These models, derived from generalized linear latent and mixed models, also called GLLAMMs, use an amalgam of the characteristics related to the generalized linear mixed models, or GLMMs, as well as SEMs, in order to utilize both the measurement model and the structural model for latent variables. Furthermore, the measurement model generalizes the GLMMs in order to merge factor structures as well as random intercepts and coefficients. This is similar to GLMMs in that the data can contain a subjective number of levels and is often very uneven, or biased with a mixture of numbers from lower-level units in the higher-level units and missing data.

As such, a diverse spectrum of SEMs can be used as the modeling methodology including: counts, ordered and unordered categorical responses, and mixed-type responses that incorporate both of these elements. According to this design, the structural models are very analogous to the structural components of the structural equation models, excluding the fact that it can include latent and observed variables are found to be fluctuating at various levels. This can be seen in the case of factors of random coefficients or unit-level latent variables, which can be regressed onto clusterlevel latent variables. As such, the Maximum likelihood estimation, as well as the empirical Bayes latent score prediction, can both be implemented when using adaptive quadrature in GLLAMM[25].

In a study relevant to these findings, Song and Lee[26] created a strong body of evidence surrounding the multiple group nonlinear structural equation models. The created models, focused on those equations with missing continuous and dichotomous data. They used models that work with data that is missing at random using the maximum likelihood approach. This demonstrated that the recently established systems for estimating and actively creating comparative models by a simulation study, as well as real data applications were valid. As a result, Koh and Zumbo[27] used multiple group confirmatory factor analysis to study the measurement of invariance in mixed item format data which is ordinal in nature. They more specifically used this methodology to research the empirical Type I error rates generated when implementing the maximum likelihood that are ordinal in nature and investigated the empirical Type I error rates of using estimation method and Pearson covariance matrix of full and strong measurement invariance hypotheses.

Also, Song et al.[28] created a two-level SEM in order to analyze multi-variate longitudinal responses, using ordered categorical and mixed continuous variables. Song determined that the first-level model serves as a definition for the measurements taken at each established moment, or time point, for investigating the change in individual characteristics that occurred during the time specified. Thus the second level allows individuals to assess the groups' characteristics as they become invariant with time. More specifically, this work demonstrates that the model accommodates covariates, missing data, and nonlinear terms of the latent variables. As such, the maximum likelihood, or the ML method, is demonstrated effective for the estimation of parameters and model comparison. So, the ML estimation is satisfactory, and the final results of a simulation study indicate that it is a good performance of this method.

Structural Equation Models for Ordered Categorical and Dichotomous Variables

The structural equation models that implement either the combination of ordered categorical, dichotomous, or continuous outcome variables are identified as the continuous/ categorical variable methodology, or CVM. When using this method, the bivariate associations between the observed variables are estimated using polychoric correlations. As such, the CMV is assumed to be both normal, and a continuous process which motivates the observed variables. This model is estimated using the ELS method, and by implementing the values of the corrected test statistics which are provided[29].

More specifically, according to the work of Muth˙enand Asparouhov[30], the CVM approach can be used in M-plus for each of the observed ordinal indictors associated with an specific underlying latent response variable. In this case, the underlying amount of a continuous and normally distributed trait or characteristic must respond in a certain category of the corresponding observed ordinal data, so that the observed indicator is dichotomous. Thus, revealing all items with a true-false response format. This insinuates that the threshold is the point for the latent response variable at which only one solution is generated. So, in this case, if the answer derived is a "true" it represents the fact that the threshold was exceeded, but if the answer is determined to be a "false" it simply represents that the threshold was not unexceeded [31].

This is significant because dichotomous items have a single threshold. However, the total number of thresholds for any item is related to three response categories, depending on the number of the category. So, each latent response variable is characterized according the continuous indicator of the underlying substantive factor, which relates to a specific hypothetical construct. The result is that the data matrix analyzed is an asymptotic correlation matrix of the latent response variables.

In contrast, for dichotomous indicators, the corresponding matrix will be a tetrachoric correlation matrix, representing the total correlation between two dichotomous variables. However, for items with at least three response categories, the data matrix will be an estimated polychoric correlation matrix, representing the correlation between two ordered categorical variables. It should also be considered that the work of Skrondal and Rabe-Hesketh[32] implemented categorical variables in the SEM model. Traditionally, structural equation models have been simplified, or generalized, in order to accommodate a variety of response type, including noncontinuous, ordinal and nominal variables, dichotomous variables, durations, and counts.

However, Montfort et al.[33] posited that similar SEMs with non-normal variables can be altered using transformation, if in the context of maximum likelihood, and when the least squares estimate is a generalization of the model parameter. In a complimentary study, Deniz et al.[34] introduced the concept which provides a viable, if nontraditional alternative technique for analyzing dichotomous, categorical, and mixed sets of data via SEMs. This method was capable of avoiding many of the downfalls of other similar models currently in use. More specifically, their approach implemented the Gifi system as a foundation. This system relies on the optimal scaling methodology in order to create meaningful numerical data about the total observed categorical variables.

As a result of these calculations, information related to the observed variable is retained with the terms of the quantified variables, or more exactly, the Gifi system changes categorical data into continuous data without damaging or disregarding the data's scaling properties. That data is preserved as part of the transformed nonlinear continuous data space within the Gifi, which also has the benefit of being invertible. Overall, this method gives legitimacy to the presumptions made about distribution, as the result of using the multivariate typicality inside an SEM.

Also of interest is Kim and Yoon[35] more recent study which compared the application of both confirmatory factor analysis and item response theory as it relates to categorical and ordinal data sets. The two major approaches Kim and Yoon considered for testing the measure of invariance for ordinal data, included: multiple group categorical confirmatory factor analysis, often simply called MCCDA, and item response theory, specifically.

This theory was ideologically different than traditional ordinary line factor analysis, in that MCCDA is capable of providing an approximate model of the ordered categorical measure without the need for a threshold structure. For example, a simulation study was carried out with the purpose of comparing these two approaches, in order to determine which had the greater power to detect the lack of invariance across a data set, or group. It was determined that both MCCFA and item response theory have a marked ability to identify the non-invariant item, assuming that the DIF or differential item function, was large enough. However, it should be noted that in spite of this effectiveness, both methodologies also yielded relatively significant false positive rates; an effect that can be minimized in the MCCDA approach if the critical values are adjusted to stabilize performance. To expand on this concept, consider the alternative model fit indexes for the MCCFA.

Under close scrutiny, this approach has been demonstrated to be reliable in detecting DIF in general, howeverPoon and Wang[36] have done extensive work on latent variable models that give insightful details on the issue. They specifically found that a more general latent variable model can be used in cases which have ordinal categorical variables, and in which both the covariates and the responses are ordinal in nature, in order to evaluate the effect of covariates on responses within the model, and in order to model the covariance structure of the responses as well. This can be accomplished via a totally Bayesian approach. More specifically, a Gibbs sampler can be applied to simulate the latent variables and parameters' joint posterior distribution. This also applies to the parameter's expansion and reparameterization, and can speed up the convergence procedure in the problem. Further, Rhemtulla et al.[37] compared the results between robust continuous SEMS and categorical SEMS in determining the estimation method most effective in what would otherwise be considered sub-optimal condition, in which categorical variables were treated as continuous. They found that in most conditions moth methods are equally acceptable for use.

The Bayesian Analysis of Structural Equation Models

Taking time to consider the literature available on the Bayesian approach is essential to this study, overall, because the approach is becoming increasingly popular for use in the SEM related calculation. This is because the Bayesian approach is effective in managing the issues that arise when working with complex SEMs and multifaceted data structures It treats unknown parameter vectors with in the model as random variables, and then analyzes the posterior distribution of those calculations, in order to essentially create and consider the conditional distribution of the indicated data set.

This process has been well demonstrated by Song and Lee[38], who compared the effectiveness of the Bayesian model, as opposed to and maximum likelihood approach, as it relates to the analysis of SEMs with a small sample size. They also considered the confirmatory fact analysis for comparison. As a result, the developed a Bayesian analysis of a two-level nonlinear structural equation model with both continuous and plytomous data sets. They thus dealt with both between group and within group levels, despite their unique difficulties. The found that a Markov-Chain Monte Carlo procedure, which is based on a blending of the Gibbs sampling method and the Metropolis–Hasting algorithm in order to create a joint Bayesian estimates of the related thresholds, and provide an overview of both the structural parameters and latent variables at each of the aforementioned levels.

They based their work on early mathematical models designed by Good (1953,1956,1965) which were designed to aiding in smoothing the proportions within the contingency tables. They also considered the work of Lindley (1964) which outlined a model of inference about odds ratios. Both of these approaches principally relied on conjugate beta and Dirichlet priors. Further, Altham (1969, 1971) offered that Bayesian analogs, which are based on small-sample frequentist tests for 2×2 tables using such priors, provide an alternative method for using normal priors for logits to solve these and similar problems. This allowed for greater flexibility and increased the overall scope for generalizations. Ultimately, however the availability of modern computational methods, which have come into existence since the mid-1980s, has shifted the focus to full Bayesian analyses with models for categorical data, allowing a primary emphasis on generalized linear models. This work has been largely carried out by Lee and his colleges.

In a recent study, Lee et al.[39] determined that the Bayesian approach can be used to investigate a general structural equation, and can accommodate the general nonlinear terms related to latent variables and realated covariates. Their methodology specifically generates Bayesian estimates that have the same statistical optimal properties as the related maximum likelihood estimate.

In a second, concurrent, study Song and Lee[40] generated a Bayseian analysis for latent variable models that have non-ignorable missing outcomes within their exponential family. Their method provides a complete framework for the analysis of complex non-normal medical and biological data, and proposes a Bayesian approach to the non-linear latent variables models of the same type. This is significant, for example, because the proposed methods can provide meaningful illustration of the outcomes for real data. In this case, it demonstrated the models use on data regarding the nonadherence of hypertension patients to med protocols.

In another, more expansive study by Lee and Xia[41], a more robust Bayesian method was used in relation to SEMs with missing data, over normal, independent distributions. This included multivariate t distributions, multivariate contaminated distributions, and multivariate slash distributions among others. These distributions were then used to develop a robust Bayseian approach that could be used to analyze the SEMs with both complete or missing data sets. These methods were specifically established to accommodate the creation of estimation data and model comparisons, and outcomes of the simulation study determined that they were effective in revealing the characteristics of estimation for the data involved. The specific methods employed were demonstrated using the real data sets for diabetes patients, demonstrating its specific relevance in the medical field.

Song and Lee[42] similarly presented the idea that a two level SEM model could be used to analyze hierarchical data that has missing entries in the data set. They also described a Bayesian approach to estimating and generating a model comparison. This compliments our own study describing how to use Win BUGS software in order to generate these forms of solutions with greater convenience. Song and Lee's purposed methods can be better illustrated through a simulation study, which allows for real application of these concepts as they relate to organizational management of research, especially with regard to the study of relationships between the latent constructions and descriptive variables within the study. In variation, several prominent researchers have offered alternative ways to use this basic premise.

Lee and Song[43] offered an alternative definition of Bayesian estimation and model comparison that focused on the integrated structural equation model. Further, Song et al.[44] recommended that Bayesian semiparametic analysis could be used for structural Equation Models that contain mixed continuous or unordered categorical variables. Cai and Song[45] found that Bayesian analysis of mixtures could be used in SEM's wow non-ignorable missing data with great success. Similarly, Stokes-Riner [46] implemented residual diagnostics as a way for Bayesian SEMs to be considered. Finally, Asparouhov and Muth˙en[47] described how to use Mplus to derive multiple significant modeling possibilities. These concepts, together offer a variety of new possibilities for use, but also pose a distinct set of challenges because the use of the method can be quite complex.

In further work on the topic, Yang and Dunson[48] introduced the idea of a broader range of Bayesian structural equation models, all of which can be used with manifest continuous variables and mixed categorical variables, while still accounting for latent variables which have unknown distribution. Similarly Song et al.[49] stated that a Bayesian approach could be taken for longitudinal analysis of SEMs. This is a two-layered approach, that accommodates missing data, and which was developed for the estimation of parameters and the generation of model comparison.

Song et al.[50] similarly investigated the concept of using a Bayesian model comparison statistic, more specifically the L measure to find both semi-parametric and parametric SEMs. This approach has a wide set of applications, with real data, and allows for model comparison. Wang and Fan[51] proposed a similar SEM, that contained a time feature series not common in other methodologies. Their approach, which was Bayesian in form, can be used to solve models by using the Markov Chain Monte Carlo method. This generates both inferences and production with the proposed time series SEM and reveals otherwise easily overlooked relationships.

Song et al.[52] similarly established an approach for generalized random coefficient SEM for generating adjacent time effect related data. Furthermore, Song and Lee[53] created a tutorial exposition that revealed how the Bayesian model, as an SEM analysis tool, could be used as a regression method with both latent and observed variables. The Bayesian approach has greater flexibility when handling complex data than other, more traditional approaches, including those based on computer software model creation.

In accordance with these findings, Chen et al.[54] created a Bayesian diagnostic procedure, used to transform structural equation models. The ability to transform structural equation models is useful, because it gives the statistician the ability to deal with non-normality of data that has several dimensions, and can demonstrate a greater degree of interrelatedness between latent variables. Additionally, the work of Yanuar and his colleges[55] instructed that Bayesian SEMS demonstrate the highest level of efficacy for creating models in the healthcare field. Yanuar et al.[55] primarily demonstrated how to model the health index based on the classical structured SEM, using a robust least squares approach and the Gibbs sampler algorithm in generating models and model data.

Using Bayesian Analysis to Investigate Structural Equation Models, specifically including those with Dichotomous Variables

These variables occur when a respondent must give a yes or no answer to a question that asks about the presence, or current experience with a specific body of symptoms. For example. "Are you feeling better today" yields a dichotomous variable. This results in a normal numerical value which is assigned to the variable, using a numbered order. For example yes would be equal to 0, while no would be equal to 1. In order to example the dichotomous data, the basic concept that SEM is a data set that comes is generated from a continuous normal distribution, which is blatantly disrupted but rigorous, which makes dichotomous nature very necessary. When looking into SEMs with dichotomous variables this offers a comparable model to those devised using categorical variable inquiry. For example, in education related research it is often necessary to explore the exclusivity of intrinsic latent factors, and related to a specific set of test points. This can be accomplished with item factor analysis of the significant model that is directly relatable to the factor structure it is based on[56,57].

There are a number of substantial, but closely related alternatives to this methodology. One of these options is to operate from the alternative focal point of analysis, which is driven by the concept that the correlated dichotomous data occurs frequently within the medical and biological fields of study in a variety of forms. Another alternative is the multivariate probit model, which is a very popular option for modeling data in biostatics. When this model is used, its data is recorded in terms of the correlated multivariate normal distribution for the underlying latent variables, which can be established as discrete variables according to the threshold requirement.

More specifically, the multivariate prohibit model which is founded on the principal structure, manages the major difficulty in analyzing and evaluating the multivariate normal probabilities created by variables that are dichotomous. This model also requires that the observation based simulation created from a multivariate truncated normal distribution have an arbitrary covariance matrix.

The potential drawback to these models, however, is that despite the ability to computer generate statistical models, the computational work needed to complete such methods is significant and time consuming. As such, this study will work to develop an approach for how to simplify the computation and reduce the total calculation by employing an SEM approach. When considering similar means to solve problems with dichotomous data, Lee and Song[57] proposed a generally simplified method using the Bayesian SEM model. This method employed unobserved continuous measurements which were underlying to the dichotomous data, and implemented an algorithm which was based on the Gibbs model in order to draw the parameter values and to compute the hypothetically missing data from the joint posterior distribution.

As an extension of this work Song and Lee[58] presented evidence that the Bayesian analysis of SEMs with nonlinear covariates, and also with latent variables, could be used to formulate a nonlinear SEM model. This model is capable of accommodating covariates within the measurement equation as well as nonlinear terms of the covariate and exogenous latent variables within the structural equations, and additionally align with Bayesian analysis of SEMs with latent variables and nonlinear covariates. More specifically, this work investigates how to formulate a nonlinear SEM that is capable of managing the measurement equations' covariates, such that the nonlinear terms of covariates and exogenous latent variables in the equation are accounted for. The covariates can be derived from both continuous and discrete distributions. Furthermore, a Bayesian approach has been developed that analyzes these models, and gives weight to each proposal. The Markov Chain Mote Carlo model provides a method for generating Bayesian estimates and the standard error estimates that are related to that model, including the highest posterior density intervals, the PP p-value and other statistical measures of interest. Also of interest is the method developed by Lee et al. [59] most recently, which suggests that implementation of the Bayesian approach for nonlinear SEMs provides a method for dealing with dichotomous variables, which implements the logit and probit link. Lee further defined how incorrectly managing binary data could produce misleading results in such models.

Using the Bayesian Analysis to Analyze SEMs with Ordered Categorical Variables

When calculating data for the social, medical, and behavioral sciences, distinctive difficulties are often encountered, specifically because their data is derived from ordered categorical variables. A common example of this issue occurs during drug studies, when patients are told to describe the effectiveness of the drug using a graduated scale that ranges from "getting worse" to "getting better." This information is categorical, and ordered, but does not provide quantitative data. As a result, many try to analyze similar ordered categorical data by assigning each variable an integer, and then treating it as continuous data from a normal distribution. This is not a significant barrier if the information is displayed in a histogram, and if the observations are generally symmetric and aligned with the normal curve, which places the most common occurances, or results that occur with the most frequency, at the center, however, in reality most data sets demonstrate a skew, or a bimodal distribution, as such this approach to ordered categorical data can create a bias in the data, or a misrepresentation of their findings. As such, a more accurate approach for measuring the discrete data is to treat each as a reflection that occurs from the latent continuous normal distribution, according to the conditions of the threshold[56].

In support of this method, Cai et al.[60] explained that the Bayesian analysis of nonlinear SEMs that contain mixed continuous, ordered or unordered categorical and nonignorable data, and which may contain missing data sets, has been widely used to examine the interaction between latent and observe variables in medical, social, and psychological research. This approach is motivated by an understanding of the way that missing data and correlated discrete variables are often encountered in these practical applications of the theoretical methodology. In these cases nonlinear structural equation models, which specifically accommodate covariates, discrete variables, mixed continuous variables, and nonignorable missing data, are appropriate.

The Bayesian methods that are used to estimate or create comparative models are then applicable. More specifically, Li and Yang [61] employed a Bayesian method that was criterion based to create a model selection of the SEMSS with ordered categorical data. This method is known as the Lv measure, and in the simulation study performed well in the area of model selection.

Analyzing Multiple Group Nonlinear Structural Equation Models According to the Bayesian Approach

The final section of pertinent literature discusses the significance of the Bayesian estimation as a model for comparison with regard to the multiple group nonlinear SEMs. This methodology can be used as a method implementing MCMC tools for data augmentation. This means, in theory, that multiple groups SEMs, or more specifically two-level SEMs can generated the desired outcome, when some conditional distributions are present and the Gibbs sample is used. With that said, however there are still certain constraints among the parameters which must be considered, as they are compulsory to working with certain groups. As such, it is key that the researcher pay special attention when stimulating the equivalent prior distribution. Similarly, using this approach requires some understanding of how to apply the path sampling procedure to these situations[15]. Similarly, Lee[56] used the latent continuous normal distribution with truncation to create solutions for the ordered categorical variables within the Bayesian multiple group SEM, combining this approach with the use of the Gibbs model for sampling to estimate the parameters. This, which was also reflected in the work of Lu et al.[62], allowed the Bayesian analysis to be applied to behavior finance and ultimately provided for the investigation of the relationship that exists between the identified influential factors, and the elements that impact motivation in that, and similar, real data situations.

The Bayesian modelling method has been developed for models of structural equation that are generalised semiparametric, and for the application in biomedical, psychological and behavioural studies Song et al.[63] Structural equation models (SEMs) are extensively used in the assessment of relationships between variables that are latent. Frequently, structural models of the regression type that are grounded in parametric functions will be utilised for these purposes.

However, in a great number of applications, parametric SEMs tend to be insufficient for capturing subtle functional patterns across the whole predictor variable range. The fact that SEMs that are traditional parametric are unable to work with a mixture of data types (count, continuous, unordered and ordered categorical) is a further and equally significant drawback. This study advances an SEM which is generalised and semiparametric which can work with a mixture of types of data and that can concurrently model a variety of functional relationships between variables that are latent. A particular series of functions that are unspecified and smooth are used to formulate the structural equation for the planned SEM. The approach of Bayesian P-splines, and the method of Markov Chain Monte Carlo, are advanced to approximate unknown parameters and smooth functions. Additionally, the relative advantages of a semiparametric model when compared to parametric models is examined with the use of DIC, or complete deviance information criterion, which is a Bayesian statistic for model comparisons. How this developed methodology performs is analysed through the use of a study of simulation. The method is illustrated through the application on a data set that originates from the National Longitudinal Survey of Youth.

Conclusion

The nonlinear models with nonlinear fixed covariate and latent varilables are very common in social and behavioural sciences. However, in SEM, examples that incorporate nonlinear terms of latent variables into equations exist. In this paper, a Bayesian approach is surveyed for analysing multiple group nonlinear models with nonlinear fixed covariate and latent variables for ordered categorical and dichotomous variables. In addition to point estimation, we provide statistical methods to obtain standard deviation estimates, and model comparisons using the Deviance Information Criterion (DIC). Owing to the complexity of the proposed model, as we have seen, difficulties arising from the nonlinear causal relationships among the nonlinear fixed covariate and latent variables, the discrete nature of ordered categorical and dichotomous variables are alleviated by data augmentation with some MCMC methods. More specifically, the basic idea of our development is inspired by the following common strategy from recent work in statistical computing that formulate the underlying complicated problem so that when augmenting the real observed data with the hypothetical missing data, the analysis would be relatively easy with the complete data. This strategy is very powerful and can be applied to other more complex models. In the future directions, we suggested using hidden continuous normal distribution (interval censoring and interval truncation) and hidden continuous normal distribution (right and left censoring, right and left truncation) to solve the problem of ordered categorical and dichotomous data in Bayesian multiple group structural equation models with nonlinear covariate and latent variables for ordered categorical and dichotomous data.

References

1. Khine M.S., Application of Structural Equation Modeling in Educational Research and Practice, *Springer* ISBN 946209330X, **(2013)**

2. Byrne B.M., Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming. New York, NY [u.a.]: Routledge, ISBN 9780805863727 0805863729 9780805863734 0805863737, **(2010)**

3. Wang J. and Wang X., Structural Equation Modeling: Applications Using Mplus. John Wiley and Sons, ISBN 1118356314, **(2012)**

4. Muthén L.K. and Muthén B.O. Mplus: statistical analysis with latent variables; user's guide; [Version 7], Muthén and Muthén **(2012)**

5. Lee S.Y., Song X.Y., Skevington S. and Hao Y.T., Application of structural equation models to quality of life, *Structural equation modeling,* **12(3),** 435–453, ISSN 1070-5511, **(2005)**

6. Song X. and Lee S.Y., A multivariate probit latent variable model for analyzing dichotomous responses, Statistica Sinica. **15(3),** 645. ISSN 1017-0405, **(2005)**

7. Fox J., Teacher's Corner: Structural Equation Modeling with the sem Package in R. *Structural Equation Modeling: A Multidisciplinary Journal,* **13(3),** 465–486. ISSN 1070-5511 1532-8007, **(2006)**

8. Asparouhov T. and Muth˙en B., Exploratory structural equation modeling, *Structural Equation Modeling: A Multidisciplinary Journal,* **16(3),** 397–438, **(2009)**

9. Yang Y. and Green S.B., A Note on Structural Equation Modeling Estimates of Reliability, *Structural Equation Modeling: A Multidisciplinary Journal,* **17(1),** 66–81. ISSN 1070-5511 1532-8007, **(2010)**

10. Iacobucci D., Structural Equations Modeling: Fit Indices, Sample Size, and Advanced Topics, *Journal of Consumer Psychology,* **20(1),** 90–98. ISSN 10577408, **(2010)**

11. Markus K.A., Structural Equations and Causal Explanations: Some Challenges for Causal SEM, *Structural Equation Modeling: A Multidisciplinary Journal,* **17(4),** 654–676. ISSN 1070-5511 1532-8007, **(2010)**

12. Hildreth L., Residual Analysis for Structural Equation Modeling. Ph.D. Thesis, **(2013)**

13. Wu J.Y. and Kwok O.M., Using SEM to Analyze Complex Survey Data: A Comparison between Design-Based Single-Level and Model-Based Multilevel Approaches, *Structural Equation Modeling: A Multidisciplinary Journal,* **19(1),** 16–35. ISSN 1070-5511 1532-8007, **(2012)**

14. Paul W.L. and Anderson M.J. Causal modeling with multivariate species data, *Journal of Experimental Marine Biology and Ecology,* **448,** 72–84. ISSN 0022-0981, **(2013)**

15. Lee S.Y. and Song X.Y., Basic and Advanced Structural Equation Models for Medical and Behavioural Sciences, Hoboken: Wiley, ISBN 9780470669525 0470669527, **(2012)**

16. Lee S.Y. and Song X.Y., Model Comparison of Nonlinear Structural Equation Models with Fixed Covariates, Psychometrik, **68(1),** 27–47, **(2003b)**

17. Lee S.Y. and Tang, N.S., Analysis of Nonlinear Structural Equation Models with Nonignorable Missing Covariates and Ordered Categorical Data, *Statistica Sinica,* **16(4),** 1117. ISSN 1017-0405, **(2006)**

18. Lee S.Y., Song X.Y., Cai J.H., So W.Y., Ma C.W. and Chan C.N.J. Non-linear structural equation models with correlated continuous and discrete data, *British Journal of Mathematical and Statistical Psychology,* **62(2),** 327-347, **(2009)**

19. Lee S.Y. and Song X.Y., Maximum Likelihood Analysis of a Two-Level Nonlinear Structural Equation Model With Fixed Covariates, *Journal of Educational and Behavioral Statistics,* **30(1),** 1–26. ISSN 1076-9986, **(2005)**

20. Henseler J. and Chin W.W., A Comparison of Approaches for the Analysis of Interaction Effects Between Latent Variables Using Partial Least Squares Path Modeling, *Structural Equation Modeling: A Multidisciplinary Journal,* **17(1),** 82– 109. ISSN 1070-5511 1532-8007, **(2010)**

21. Wen Z., Marsh H.W. and Hau K.T., Structural Equation Models of Latent Interactions: An Appropriate Standardized Solution and Its Scale-Free Properties, *Structural Equation Modeling: A Multidisciplinary Journal,* **17(1),** 1–22. ISSN 1070-5511 1532-8007, **(2010)**

22. Codd C.L., Nonlinear Structural Equation Models: Estimation and Applications, Ph.D. Thesis, The Ohio State University, **(2011)**

23. Pek J., Losardo D. and Bauer D.J., Confidence Intervals for a Semiparametric Approach to Modeling Nonlinear Relations among Latent Variables, *Structural Equation Modeling: A Multidisciplinary Journal,* **18(4),** 537–553. ISSN 1070-55111532-8007, **(2011)**

24. Lee S.Y., Bayesian Analysis of Nonlinear Structural Equation Models with Nonignorable Missing Data, *Psychometrika,* **71(3),** 541–564. ISSN 0033-3123 1860-0980, **(2006)**

25. Rabe-Hesketh S., Skrondal A. and Pickles A., Generalized multilevel structural equation modeling, *Psychometrika,* **69(2),** 167–190. ISSN 0033-3123, **(2004)**

26. Song X.Y. and Lee S.Y., A Maximum Likelihood Approach for Multisample Nonlinear Structural Equation Models with Missing Continuous and Dichotomous Data, *Structural Equation Modeling,* **13(3),** 325–351. ISSN 1070- 5511, **(2006b)**

27. Koh K.H. and Zumbo B.D., Multi-Group Confirmatory Factor Analysis for Testing Measurement Invariance in Mixed Item Format Data, *Journal of Modern Applied Statistical Methods,* **7(2),** 12. ISSN 1538-9472, **(2008)**

28. Song X.Y., Lee S.Y. and Hser Y., A two-level structural equation model approach for analyzing multivariate longitudinal responses, *Statistics in medicine,* **27(16),** 3017–3041. ISSN 1097-0258, **(2008)**

29. Kline R.B., Principles and Practice of Structural Equation Modeling. New York: Guilford Publications, ISBN 9781606238769 1606238760, **(2011)**

30. Muth˙en B. and Asparouhov T., Latent Variable Analysis with Categorical Outcomes: Multiple-Group and Growth Modeling in Mplus, Mplus Web Notes, **4(5),** 1–22 **(2002)**

31. Brown T.A., Confirmatory Factor Analysis for Applied Research, **(2006)**

32. Skrondal A. and Rabe-Hesketh S., Structural equation modeling: categorical variables, Wiley Online Library, **(2005)**

33. Montfort K.V., Mooijaart A. and Meijerink F., Estimating Structural Equation Models with Non-Normal Variables by Using Transformations, *Statistica Neerlandica,* **63(2),** 213–226. ISSN 00390402 14679574, **(2009)**

34. Deniz E., Bozdogan H. and Katragadda S., Structural equation modeling (SEM) of categorical and mixed-data using the Novel Gifi transformations and information complexity (ICOMP) criterion, *Journal of the School of*

Business Administration, Istanbul University, **40(1),** 86–123. ISSN 1303-1732, **(2011)**

35. Kim E.S. and Yoon M., Testing Measurement Invariance: A Comparison of Multiple-Group Categorical CFA and IRT, *Structural Equation Modeling: A Multidisciplinary Journal,* **18(2),** 212–228. ISSN 1070-5511 1532-8007, **(2011)**

36. Poon W.Y. and Wang H.B. Latent variable models with ordinal categorical covariates, Statistics and Computing, **22(5),** 1135–1154. ISSN 0960-3174, **(2012)**

37. Rhemtulla M., Brosseau Liard P.E. and Savalei V., 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. ISSN 1939-1463, **(2012)**

38. Song X.Y. and Lee S.Y., 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, **(2004)**

39. Lee S.Y., Song X.Y. and Tang N.S., Bayesian Methods for Analyzing Structural Equation Models with Covariates, Interaction, and Quadratic Latent Variables. Structural Equation Modeling: A Multidisciplinary Journal, **14(3),** 404– 434. ISSN 1070-5511 1532-8007, **(2007)**

40. Song X.Y. and Lee S.Y., Bayesian analysis of latent variable models with non-ignorable missing outcomes from exponential family, *Statistics in Medicine,* **26(3),** 681–693, **(2007)**

41. Lee S.Y. and Xia Y.M., A Robust Bayesian Approach for Structural Equation Models with Missing Data, *Psychometrika,* **73(3),** 343–364, ISSN 0033-3123 1860-0980, **(2008)**

42. Song, X.Y. and Lee S.Y., A Bayesian Approach for Analyzing Hierarchical Data with Missing Outcomes Through Structural Equation Models, *Structural Equation Modeling,* **15(2),** 272–300. ISSN 1070-5511, **(2008)**

43. Lee, S.Y. and Song X.Y., On Bayesian Estimation and Model Comparison of an Integrated Structural Equation Model, *Computational Statistics and Data Analysis*, **52(10),** 4814–4827. ISSN 01679473, **(2008)**

44. Song X.Y., Xia Y.M. and Lee S.Y., Bayesian semi Parametric analysis of structural equation models with mixed continuous and unordered categorical variables, *Statistics in Medicine,* **28(17),** 2253–2276, **(2009)**

45. Cai J.H. and Song X.Y., Bayesian analysis of mixtures in structural equation models with non-ignorable missing data, *British Journal of Mathematical and Statistical Psychology,* **63(3),** 491-508, **(2010)**

46. Stokes Riner A., Residual Diagnostic Methods for Bayesian Structural Equation Models. Ph.D. Thesis. University of Rochester, **(2009)**

47. Asparouhov T. and Muth˙en B., Bayesian analysis of latent variable models using Mplus, Retrieved June, **17,** 2014. **(2010)**

48. Yang M. and Dunson D.B., Bayesian Semiparametric Structural Equation Models with Latent Variables, *Psychometrika,* **75(4),** 675–693. ISSN 0033-3123 1860-0980, **(2010)**

49. Song X.Y., Lu Z.H., Hser Y.I. and Lee S.Y., A Bayesian Approach for Analyzing Longitudinal Structural Equation Models, *Structural Equation Modeling: A Multidisciplinary Journal,* **18(2),** 183–194. ISSN 1070-5511 1532-8007.**(2011a)**

50. Song X.Y., Xia Y.M., Pan J.H. and Lee S.Y., Model Comparison of Bayesian Semiparametric and Parametric Structural Equation Models, *Structural Equation Modeling: A Multidisciplinary Journal,* **18(1),** 55–72. ISSN 1070-5511 1532-8007. **(2011b)**

51. Wang Y.F. and Fan T.H., A Bayesian analysis on time series structural equation models, *Journal of Statistical Planning and Inference,* **141(6),** 2071–2078. ISSN 0378-3758. **(2011)**

52. Song X.Y., Tang N.S. and Chow S.M., A Bayesian approach for generalized random Coefficient structural equation Models for Longitudinal data with Adjacent Time effects, *Computational Statistics and Data Analysis,* **56(12),** 4190– 4203. ISSN 0167-9473, **(2012)**

53. Song X.Y. and Lee S.Y., A Tutorial on the Bayesian Approach for Analyzing Structural Equation Models, *Journal of Mathematical Psychology,* **56(3),** 135–148. ISSN 00222496, **(2012)**

54. Chen J., Liu P. and Song, X.Y., Bayesian Diagnostics of Transformation Structural Equation Models, *Computational Statistics and Data Analysis,* **68,** 111– 128. ISSN 0167-9473, **(2013)**

55. Yanuar F., Ibrahim K. and Jemain A.A., Bayesian Structural Equation Modeling for the Health Index, *Journal of Applied Statistics,* **40(6),** 1254–1269, ISSN 0266-4763 1360-0532, **(2013)**

56. Lee S.Y., Structural Equation Modeling: A Bayesian Approach, Chichester, England; Hoboken, NJ: Wiley. ISBN 9780470024232 0470024232, **(2007)**

57. Lee S.Y. and Song X.Y., Bayesian Analysis of Structural Equation Models with Dichotomous Variables, *Statistics in Medicine,* **22(19),** 3073–3088. ISSN 1097- 0258, **(2003a)**

58. Song X.Y. and Lee S.Y., Bayesian analysis of structural equation models with nonlinear covariates and latent

variables, *Multivariate Behavioral Research,* **41(3),** 337–365. ISSN 0027-3171, **(2006a)**

59. Lee S.Y., Song, X.Y. and Cai J.H., A Bayesian Approach for Nonlinear Structural Equation Models with Dichotomous Variables Using Logit and Probit Links, *Structural Equation Modeling,* **17(2),** 280–302. ISSN 1070-5511, (2010)

60. Cai, J.-H., Song, X.-Y. and Lee, S.-Y. Bayesian Analysis of Nonlinear Structural Equation Models with Mixed Continuous, Ordered and Unordered Categorical and Nonignorable Missing Data, *Statistics and its Interface,* **1,** 99–114, **(2008)**

61. Li, Y. and Yang, A. A Bayesian Criterion-Based Statistic for Model Selection of Structural Equation Models with Ordered Categorical Data, *International Journal of Modeling and Optimization,* **1(2), (2011)**

62. Lu, B., Song, X.-Y. and Li, X.-D. Bayesian Analysis of Multi-Group Nonlinear Structural Equation Models with Application to Behavioral Finance, *Quantitative Finance*, **12(3),** 477–488, ISSN 1469-7688 1469-7696, **(2012)**

63. Song X.Y., Lu Z.H., Cai J.H. and Ip E. H.S., A Bayesian Modeling Approach for Generalized Semiparametric Structural Equation Models, *Psychometrika,* **78(4),** 624–647. ISSN 0033-3123, **(2013)**