GJPAM. The use of uninfo & info Baye SEM. Ferra in GJPAM 2015.pdf

by Ferra Yanuar

Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 5 (2015), pp. 3259-3264

http://www.ripublication.com

The Use of Uniformative and Informative Prior Distribution in Bayesian SEM

Ferra Yanuar

Department of Mathematics, Faculty of Mathematics and Natural Sciences
ferrayanuar@yahoo.co.id

Abstract

This study applies Bayesian approach to the construction of health status model. Bayesian combines prior distributions with the data likelihood to form posterior distributions to estimate the parameters. An algorithm based on the Gibbs sampler is applied for drawing the parameters values from the joint posterior distributions. Some criteria are used to test the goodness of fit of the posited model. Uninformative and informative prior are used in Bayesian analyses. Since health status model involves observed and unobserved variables simultaneously, Bayesian analysis is then combined with structural equation modeling (SEM) approach in fitting the hypothesis model to the data. The main purpose of this study is to demonstrate the application of uninformative and informative prior in Bayesian SEM to construct the health status model of an individual. Two real data sets are considered in this study. First, data set uses uninformative prior in parameter estimation, which then be adopted as informative prior for the second data set. This study proves that model uses informative prior results better estimation on parameter model than uninformative. This study also informs that socio-demography and lifestyle have greater effect to the health condition of an individual than to mental health.

Keywords: Bayesian, health status, prior distribution, informative prior, structural equation modeling.

Introduction

In the survey where data arising in the fields of sociology, psychology, health or economics, the data often use the predictor and outcome variables which can both be either unobserved (latent) or observed. Such modeling requires the implementation of structural equations modeling (SEM) as it is a powerful multivariate regression technique to handle scenarios in modeling for exploring relationships among latent and observed variables.

Many researchers, such as Lee *et al.* [1] andYanuar*et al.* [2] proposed to use Bayesian approach in SEM to analyze non-normal data. Bayesian analysis uses the Gibbs sampler [3] to handle the complexities of the model and the multi-dimensional integrals problems in order to obtain samples of arbitrary size. From these samples, the user can compute the point estimates, standard deviation and interval estimates. The basic attractive feature of Bayesian approach is to allow the user to use the prior information for updating the current information regarding the parameter of interest. Bayesian combines prior distributions for parameters with the data likelihood to form posterior distributions for the parameter estimates. The priors can be non-informative or informative. Informative prior is obtained from previous studies or based on experts opinion.

In this article, our main purpose is to demonstrate the use of uninformative and informative prior in Bayesiananalysis to construct the health status model of an individual living in urban area in West Sumatera Indonesia. We will identify the relationships among latent variables, such as health status, socio-demography, lifestyle and mental health. Using two set data considered, the first group uses uninformative prior for estimating the parameter model of interest. The estimated parameter obtained from the first model then are adopted as informative prior on the second group. Considerable issues of this study remain relatively unexplored.

Data and Methods

The Bayesian SEM was employed totwo data sets, i.e., the Primary Health Research in West Sumatera that conducted in the year 2007 and 2013. The survey was done by the Centre of Research and Development, Department of Health, Republic of Indonesia. The data survey in the year 2007 consist of 3677 respondents who provided complete information, meanwhile the data survey in the year 2013 involved 2651 respondents to be analyzed in this study.

The indicators for assessing the socio-demographic factor were education leveland age group [3, 4, 5, 6, 7]. The lifestyle's indicators were physical exercise and consump fiber. Meanwhile the indicators for assesing mental health were stress level and emotional problem. Health status was measured by blood pressure, general health condition and body mass index (BMI). All indicator varibles are ordinal types. In this study, as based on the literature, it was assumed that socio-demography, lifestyle, and mental health could give affect to the health status.

Both data set were modeled using Bayesian SEM analysis, which involves two major components; namely, the measurement equation and the structural equation. In classical SEM analysis, parameter estimation is based on Maximum Likelihood approach. Meanwhile in Bayesian SEM, it use prior distribution for obtaining parameter estimate. The data distribution y given the unobserved parameter θ, $p(y|\theta)$ and the prior about θ, $p(\theta)$ are spesified in order to obtain the posterior distribution. In Bayesian estimations, parameters are considered as random with prior distribution and

a prior density function [8, 9]. The parameter estimated are drawn by calculating the mean of this joint posterior distribution by applying the Gibbs sampler techniques.
In this study, we apply the uninformative prior and informative prior to estimate the parameter model. This study applies the same process as Yanuar*et al.* done to get the uninformative prior distribution, which is applied to the Primary Health research data in the year 2007. The parameter estimated obtained then are used as informative prior to model the Health Status using Primary Health Research data in the year 2013. In order to identify the informative prior distribution for the free parameter in factor loading $\boldsymbol{\Lambda}$**,** written as $\boldsymbol{\Lambda}^{(info)}$, we suggest to obtain the information from previous model and defined as follows:

$$\boldsymbol{\Lambda}^{(info)} \sim N\left(\widehat{\boldsymbol{\Lambda}}^{(uninfo)}, \widehat{Var}\left(\widehat{\boldsymbol{\Lambda}}^{(uninfo)}\right)\right) \quad (1)$$

where$\widehat{\boldsymbol{\Lambda}}^{(uninfo)}$ and $\widehat{Var}\left(\widehat{\boldsymbol{\Lambda}}^{(uninfo)}\right)$ respectively are posterior mean and variance of factor loadings in measurement equation undertaken from estimation of the first data set (Primary Health Research data in the year 2007).
The convergence of parameter is diagnosed by trace plots [8]and Monte Carlo error (an estimate of the difference between the mean of the sample values and the true posterior mean) for all the parameters is less than 5% of the sample standard deviation [9]. The accuracy of posterior estimates is inspected using Gelman-Rubin convergence statistics [3, 10].

Results

We use winBUGS version 1.4 [10] which allowed the hierarchical structure through selecting the prior information for parameters involved in the hypothesis model. Uninformative Bayesian SEM then is applied to the Primary Health Research 2007 data. Bayesian analyses are conducted using the conjugate prior distributions.Then the estimated of coefficient regression obtained from uninformative prior are used as informative prior for the second data set, Primary Health Research 2013. The results of estimation for coefficient regression in structural equation and the estimation for coefficient regression in measurement equation are provided in Table 1.

Table 1. Factor Loadings Using Uninformative and Informative Prior

Items	**Estimate of Factor Loadings**	
	UninfoPrior(SE)	**Info Prior (SE)**
Socio-demography → Health Status	-0.393 (0.046)*	-0.511 (0.031)*
Lifestyle → Health Status	0.217 (0.052)*	0.324 (0.021)*
Mental Health → Health Status	0.360 (0.064)*	0.489 (0.034)*
Socio Demography($\boldsymbol{\xi_1}$) **:**		
Age-group	0.567*	0.451*
Educational level	-0.203*	-0.301*
Lifestyle($\boldsymbol{\xi_2}$):		
Physical activity	0.302*	0.491*
Consump Fiber	0.019	0.021

Mental Health(ξ_3) :		
Stress level	0.376*	0.419*
Emotional Problem	0.799*	0.656*
Health status (η) :		
Blood Pressure	0.456*	0.342*
General health condition	0.879*	0.812*
Body Mass Index	0.287*	0.301*

SE standard error, *Significant at 5% level

Based on test of convergence for both proposed model, it is obtained that the proposed modelare adequate and can be accepted. This study also proved that model uses informative prior results better estimation on parameter model than uninformative.

Based on test of convergence, it is obtained that the proposed modelare adequate and can be accepted. This study also proved that model uses informative prior results better estimation on parameter model than uninformative since for smaller iteration the parameter estimated in informative prior have been converge compared to uninformative prior.

We obtain the estimated structural equation that addressed the relationship of health status (η)with socio-demography (ξ_1), lifestyle (ξ_2) and mental health (ξ_3) for uninformative and informative prior distribution in Bayesian SEM as follows:

η_{Uninfo} =-0.393 ξ_1+ 0.217 ξ_2+ 0.360ξ_3 (2)

η_{Info} = -0.511 ξ_1 + 0.324 ξ_2 +0.489 ξ_3 (3)

These estimated structural equation indicates that socio-demography status (ξ_1) gives negative effect on health status for both types of prior beliefs. Meanwhile lifestyle (ξ_2) and mental health $\xi_3$havepositive effect on health status. The affect of all three latent exogeneous variables on laten endogeneous variable are significant for both models.

The results of measurement equations for uninformative and informative prior distribution, as presented in Table 1. This table informs us that all factor loading estimated are significant as indicator for corresponding latent variables. One can conclude here that the older people tend to have bad health condition compared to younger. But the people who have higher education level tend to experience better health condition. It is also found that lifestyle has a direct effect on the health status for both models and this relationship is significant as well. It implies that the people will have good health condition if they follow the health lifestyle behaviors, such as often do physical activity or consump fiber such as fruits or vegetables. This study reveals that people who did not stress tend to have higher health status level than people in stress. The peoplewhohave someemotional problemstend to havepoor healththan peoplewhodo nothaveemotional problems. It is proved that to have better health condition, we must maintain our blood pressure, body mass index (BMI) and whole general health condition.

Conclusions

The main purpose of the present study is to apply the uninformative and informative prior belief in Bayesian SEM in the case of constructing a health status of an individual living in West Sumatera, Indonesia. This study involve two period times of data to be analyzed.

Bayesian combines prior distributions with the data likelihood to form posterior distributions to estimate the parameters. An algorithm based on the Gibbs sampler is applied for drawing the parameters values from the joint posterior distributions.

The informative prior for parameters are based on the estimated values resulted from previous uninformative Bayesian SEM. The health status model for both models, applied uninformative and informative prior, are almost the same (or linear). It reveals that affect of socio-demographic, mental health and lifestyle on health status is significant directly. Socio-demography and lifestyle have greater effect to the health status of an individual than to mental health. This study also proved that model uses informative prior results better estimation on parameter model than uninformative.

Based on test of the convergence of parameter model and test of accuracy of posterior estimeted, it is resulted that both proposed models are acceptable since all parameter model have been converge and accurate.

Acknowledgments

This research is supported by Hibah Bersaing, grant number Dipa-023.04.2.415061/2014 from Ministry of Higher Education, Indonesia. We thank to Hapsari from Centre of Research & Development, Department of Health, Republic of Indonesia for supplying data of Primary Health Research 2007 and 2013. We also thank to several anonymous referees for their constructive comments which have improved the final version of this paper.

References

[1] Lee SY, Song XY, Skeviton S&Hao YT, 2005, "Application of structural equation models to quality of life". *Structural equation modeling,* 12(3), 435-453.

[2] Yanuar, F., Ibrahim, K. &Jemain, A.A., 2013, "Bayesian Structural Equation ModelingFor The Health status". *Journal of Applied Statistics,* 40 (6), 1254-1269.

[3] Brooks SP, Gelman A., 1998, "Alternative methods for monitoring convergence of iteration simulations". *Journal of Computation Graph Statistics,* 7, 434-455.

[4] Cheadle A, Pearson D, Wagner E. et al., 1994, "Relationship between socioeconomic status, health status, and lifestyle practices of American Indians: Evidence from a plains reservation population". *Public Health Report,* 109(3), 405-413.

[5] Uitenbroek D.G, KerekovskaA&Festchieva N., 1996, "Health lifestyle behaviour and socio-demographic characteristics. A study of Varna, Glasgow and Edinburgh". *Social Science & Medicine,* 43 (3), 367-377.

[6] Shi L., 1998, "Socio-demographic characteristics and individual health behaviors". *Socio demographic factors and health behaviors,* 91 (10), 933-942.

[7] Boardman JD., 2004, "Stress and physical health: the role of neighborhoods as mediating and moderating mechanisms". SocSci Med. 58: 2473–83.

[8] Lee, S.Y., 2007, *Structural equation modeling: A Bayesian approach*. John Wiley & Sons, Ltd, New York.

[9] Yanuar, F., 2014, "The Estimation Process in Bayesian Structural Equation Modeling Approach". *Journal of Physics: Conference Series*, 495, 012047

[10] Spiegelhalter D, Thomas A, Best N&Lunn D., 2003, WinBUGS User Manual, Version 1.4. Available at http://www.mrc-bsu.cam.ac.uk/bugs,

GJPAM. The use of uninfo & info Baye SEM. Ferra in GJPAM 2015.pdf

ORIGINALITY REPORT

%26	%14	%22	%5
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.i-scholar.in Internet Source	%5
2	Ferra Yanuar. "On the application of structural equation modeling for the construction of a health index", Environmental Health and Preventive Medicine, 04/06/2010 Publication	%4
3	www.statmodel.com Internet Source	%3
4	Sik-Yum Lee. "Bayesian analysis of structural equation models with dichotomous variables", Statistics in Medicine, 10/15/2003 Publication	%2
5	Ferra Yanuar. "The Estimation Process in Bayesian Structural Equation Modeling Approach", Journal of Physics: Conference Series, 2014 Publication	%2

Yanuar, Ferra. "The Health Status Model in

6	Urban and Rural Society in West Sumatera, Indonesia: An Approach of Structural Equation Modeling", Indian Journal of Science and Technology, 2016. Publication	%2
7	"A Bayesian Structural Equations Models for Multilevel Data with Missing Responses and Missing Covariates", Bayesian Analysis, 2008. Publication	%2
8	www.scirj.org Internet Source	%2
9	Submitted to University of Malaya Student Paper	%2
10	www.ripublication.com Internet Source	%1
11	Adam Sagan. "Dyadic Interactions in Service Encounter: Bayesian SEM Approach", Studies in Classification Data Analysis and Knowledge Organization, 2009 Publication	<%1

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		