Tourism Management 64 (2018) 98–109

Contents lists available at ScienceDirect

Tourism Management

journal homepage: www.elsevier.com/locate/tourman

The time has come: Toward Bayesian SEM estimation in tourism research☆

A. George Assaf [a,*], Mike Tsionas [b], Haemoon Oh [c]

[a] Isenberg School of Management, University of Massachusetts-Amherst, 90 Campus Center Way, 209A Flint Lab, Amherst, MA, 01003, United States
[b] Lancaster University Management School, United Kingdom
[c] College of Hospitality, Retail and Sport Management, University of South Carolina, United States

HIGHLIGHTS

- We discuss the power of the Bayesian approach for SEM estimation.
- We compare between the Bayesian and covariance based approaches in small sample sizes.
- We discuss several SEM contexts where the Bayesian approach provides unique advantages.

ARTICLE INFO

Article history:
Received 17 June 2017
Received in revised form
14 July 2017
Accepted 29 July 2017

Keywords:
Bayesian approach
SEM
Small samples
Monte Carlo simulation

ABSTRACT

While the Bayesian SEM approach is now receiving a strong attention in the literature, tourism studies still heavily rely on the covariance-based approach for SEM estimation. In a recent special issue dedicated to the topic, Zyphur and Oswald (2013) used the term "Bayesian revolution" to describe the rapid growth of the Bayesian approach across multiple social science disciplines. The method introduces several advantages that make SEM estimation more flexible and powerful. We aim in this paper to introduce tourism researchers to the power of the Bayesian approach and discuss its unique advantages over the covariance-based approach. We provide first some foundations of Bayesian estimation and inference. We then present an illustration of the method using a tourism application. The paper also conducts a Monte Carlo simulation to illustrate the performance of the Bayesian approach in small samples and discuss several complicated SEM contexts where the Bayesian approach provides unique advantages.

1. Introduction

Over the last two decades, structural equation modelling (SEM) has become one of the most popular methodologies in tourism research. The method's popularity stems from its ability to handle complicated relationships between latent and observed variables, which are highly common in tourism research (Reisinger & Turner, 1999). While relatively a complex method, the availability of several SEM software packages (e.g. AMOS, LISREL, M*plus*) has certainly facilitated the widespread application of the method and brought it within the reach of the applied researcher (Assaf, Oh, & Tsionas, 2016). Basically, SEM consists of the "measurement equation", which is like a regression model between the latent and observed variables, and the "structural equation", which is a regression between the latent variables. With latent variables not being directly observed, one cannot use normal regression techniques to analyse the model.

A traditional approach in estimating SEM has been, "the covariance based approach", which focuses "in fitting the covariance structure of the model to the sample covariance matrix of the observed data" (Lee & Song, 2014, p. 276). Though in many situations, this estimation method works fine and produces reliable estimates (Assaf et al., 2016), there are some complicated data structure and model assumptions where the "covariance based approach" will encounter "serious difficulties and will be unable to produce correct results for statistical inferences" (Lee & Song, 2014, p. 277). As recently highlighted by Assaf et al. (2016), one of the main motivations for using the Bayesian approach for SEM estimation is its flexibility to handle many complicated models and/or data structures. Importantly, the "covariance approach" based on

☆ Note that all authors have contributed equally to the paper.

* Corresponding author.
E-mail addresses: assaf@isenberg.umass.edu (A.G. Assaf), m.tsionas@lancaster.ac.uk (M. Tsionas), oh@sc.edu (H. Oh).

http://dx.doi.org/10.1016/j.tourman.2017.07.018
0261-5177/

estimation methods such as maximum likelihood (ML) or generalized least squares (GLS) is only asymptotically correct (viz. it only works according to statistical theory with large sample). It is also "well known that the statistical properties of the estimates and the goodness-of-fit test obtained from these approaches are asymptotically true only" (Lee & Song, 2004, p. 653). Hence, using them in small samples should be done with caution.

Our aim in this paper is to provide for the first time a thorough introduction of the Bayesian approach for SEM estimation. Despite the growing popularity of the Bayesian approach in related fields such as Marketing and Management, it has yet to receive strong attention in the tourism literature (Zyphur & Oswald, 2013). Apart from its ability to handle more complicated SEM models, the Bayesian approach introduces several important advantages: 1) it allows the inclusion of prior information in the analysis; 2) it is more robust to small sample sizes, 3) it provides more reliable formal model comparison statistics, 4) it "provides a better approximation to the level of uncertainty, or, conversely, the amount of information provided by the model" (Rossi & Allenby, 2003, p. 306), and 5) it can be used with SEM models that include unobserved heterogeneity in the form of various random effects.

It is surprising that despite these advantages there are very limited Bayesian SEM studies in tourism (Assaf et al., 2016). We aim in this paper to introduce tourism researchers to the power of the Bayesian SEM approach, and discuss how the method can address some of the main limitations of the covariance-based approach. We discuss several interesting contexts where the Bayesian approach can help SEM researchers overcome complex model situations. With the method not being well established in the tourism literature, we start first with a brief overview of the Bayesian approach, demonstrating its advantages and illustrating how the results can be presented and interpreted. We then discuss the Markov Chain Monte Carlo (MCMC) technique, the most common method for Bayesian estimation. We follow this with an illustration of a Bayesian SEM estimation using the Winbugs software. We also conduct a Monte Carlo simulation to illustrate the advantages of the Bayesian approach over the covariance-based approach in small samples, using a well-established tourism model. The paper concludes with a discussion of several complicated SEM contexts where the Bayesian approach can provide unique advantages. Our main goal is to encourage the use of Bayesian methods for SEM estimation in the tourism literature.

2. Basic illustration of SEM

The basic linear SEM framework[1] consists of the following measurement and structural equations:

$$\begin{aligned} \underset{(p\times 1)}{y_i} &= \underset{(p\times m)}{\Lambda_y}\,\eta_i + \underset{(p\times 1)}{\varepsilon_i},\ \varepsilon_i \sim N_p\left(0,\ \underset{(p\times p)}{\Theta_\varepsilon}\right) \\ \underset{(q\times 1)}{x_i} &= \underset{(q\times n)}{\Lambda_x}\,\xi_i + \underset{(q\times 1)}{\delta_i},\ \delta_i \sim N_q\left(0,\ \underset{(q\times q)}{\Theta_\delta}\right) \end{aligned} \tag{1}$$

$$\begin{aligned} \underset{(m\times 1)}{\eta_i} &= \underset{(m\times m)}{B}\,\eta_i + \underset{(m\times n)}{\Gamma}\,\underset{(n\times 1)}{\xi_i} + \underset{(m\times 1)}{\zeta_i}, \\ \zeta_i &\sim N_m\left(0,\ \underset{(m\times m)}{\Psi}\right),\ \xi_i \sim N_n\left(0,\ \underset{(n\times n)}{\Phi}\right) \end{aligned} \qquad i = 1, \ldots, N, \tag{2}$$

where in (1), y_i and x_i are the observed variables which are the respective indicators of η_i, ξ_i, Λ_y. Λ_x are loading matrices and ε_i, and δ_i are random vectors of error measurements. Ψ, Φ, Θ_ε, and Θ_δ are the covariance matrices of ζ_i, ξ_i, ε_i and δ_i, respectively, usually assumed to be diagonal, and in (2), η_i is an endogenous latent vector, B and Γ are matrices of regression coefficients, ξ_i is an exogenous latent vector, and ζ_i is a random vector of error measurement.

From Bollen (1989, p. 325) we can find the implied covariance matrix of the model after collecting all unknown parameters into the vector $\theta \in \Theta \subseteq \mathbb{R}^d$, where d is the number of parameters and Θ is the parameter space. We have:

$$\Sigma(\theta) = \begin{bmatrix} \Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{bmatrix}, \tag{3}$$

where

$$\Sigma_{yy}(\theta) = \Lambda_y (I-B)^{-1}\left(\Gamma\Phi\Gamma' + \Psi\right)\left[(I-B)^{-1}\right]'\Lambda_y' + \Theta_\varepsilon, \tag{4}$$

$$\Sigma_{yx}(\theta) = \Lambda_y (I-B)^{-1}\Gamma\Phi\Lambda_x', \tag{5}$$

$$\Sigma_{xy}(\theta) = \Lambda_x \Phi\Gamma'\left[(I-B)^{-1}\right]'\Lambda_y', \tag{6}$$

$$\Sigma_{xx}(\theta) = \Lambda_x \Phi \Lambda_x' + \Theta_\delta. \tag{7}$$

Based on these expressions the maximum likelihood criterion to be maximized (Bollen, 1989, p. 335) is:

$$F_{ML}(\theta) = -\left\{\log|\Sigma(\theta)| + tr\left(S\Sigma^{-1}(\theta)\right)\right\} + \log\left|S\right| + (p+q), \tag{8}$$

where S is the empirical covariance matrix, the last two terms can be omitted and a "quick" necessary condition for identification is $d \le \frac{1}{2}(p+q)(p+q+1)$. Maximization of (8) is performed numerically in many commonly available software programs like AMOS, LISREL, Mplus etc. There are many situations where using this covariance based approach will encounter serious difficulties "for many complicated situations: for example, when deriving the covariance structure is difficult, or the data structures are complex" (Lee & Song, 2012, p. 15). Our goal here is to elaborate on the Bayesian estimation of SEM, illustrating its advantages and its reliability in small samples. We also present several complicated data generating processes or models where the Bayesian approach presents some unique advantages.

To set the framework for Bayesian SEM, we believe it is important to start first with description of the Bayesian approach. The literature currently lacks such description, not only within the context of SEM but within other modelling approaches. We focus on the basic ideas of Bayesian inference for both model estimation and model comparison.

3. Brief overview of the Bayesian approach

3.1. Basic concepts

The key difference between the "Bayesian approach" and the "sampling-theory or frequentist paradigm" is that in the latter one proceeds under the assumption that the coefficients are fixed but unknown. In the Bayesian paradigm, the data is treated as fixed and statistical uncertainty comes from the stochastic nature of the parameters. More often than not, in the frequentist paradigm, the exact finite-sample distributions of estimators of parameters are unknown and one has to resort to asymptotic approximations for them. Such approximations can range from totally invalid to hardly

[1] As most tourism researchers are now well familiar with SEM, we do not intend here to provide a detailed background of the method.