

# The Time Has Come: Toward Bayesian SEM Estimation in Tourism Research

## 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 and Turner, 1999). While relatively a complex method, the availability of several SEM software packages (e.g. AMOS, LISREL, *Mplus*) has certainly facilitated the widespread application of the method and brought it within the reach of the applied researcher (Assaf et al., 2016). Basically, SEM consists of two components. The first component, the “measurement equation”, is like a regression model between the latent and observed variables, while the second component, the “structural equation”, 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 and Song, 2014, p.276). Most popular SEM software such as AMOS and LISREL rely heavily on the covariance-based approach. 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 and 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 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). Using software packages such LISREL or AMOS on small sample sizes

should be done with caution, as “it is well known that the statistical properties of the estimates and the goodness-of-fit test obtained from these approaches are asymptotically true only” (Lee and Song, 2004, p. 653).

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 and 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 and 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<sup>1</sup> consists of the following measurement and structural equations:

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

---

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

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

where in (1),  $y_i$  and  $x_i$  are the observed variables which are the respective indicators of  $\eta_i$ ,  $\xi_i$ ,  $\Lambda_y$ ,  $\Lambda_x$  are loading matrices and  $\varepsilon_i$ , and  $\delta_i$  are random vectors of error measurements.  $\Psi$ ,  $\Phi$ ,  $\Theta_\varepsilon$ , and  $\Theta_\delta$  are the covariance matrices of  $\zeta_i$ ,  $\xi_i$ ,  $\varepsilon_i$  and  $\delta_i$ , respectively, usually assumed to be diagonal, and in (2),  $\eta_i$  is an endogenous latent vector,  $B$  and  $\Gamma$  are matrices of regression coefficients,  $\xi_i$  is an exogenous latent vector, and  $\zeta_i$  is a random vector of error measurement, and

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  $\Theta$  is the parameter space. We have:

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

where

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

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

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

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

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

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

where  $S$  is the empirical covariance matrix, the last two terms can be omitted and a “quick” necessary condition for identification is  $d \leq \frac{1}{2}(p+q)(p+q+1)$ . 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 and 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.

### 3.1. Brief Overview of the Bayesian Approach

#### 3.1.1. 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. Uncertainty is introduced into the analysis because the data is viewed as one particular realization among many so there is sampling-variation in parameter estimates. 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 acceptable. In the Bayesian paradigm, we can derive exact posterior distributions of the parameters given the data using Bayes’ theorem which combines the likelihood and the prior. The prior is indeed a distinguishing feature that quantifies *a priori* uncertainty in Bayesian analysis, and summarizes all knowledge that we have (from theory or previous studies) about the parameters before observing the data. There is no need to resort to asymptotic approximations when the data set is small and, therefore, we expect more precise statistical inferences. In addition, model selection and the whole inference apparatus become rather easy once we adopt the Bayesian approach. Of course, asymptotically, under any prior, the Bayesian posteriors converge to normal distributions with moments given by the usual ML quantities.

To better understand how Bayesian analysis works, we start first with specifying the likelihood of the data,  $L(\theta; \mathcal{D})$ , given an unobserved parameter “ $\theta$ ” and the *given* data,  $\mathcal{D}$ . In the frequentist approach,  $\theta$  is treated as unknown but fixed, while with the Bayesian approach  $\theta$  is treated as random (Kaplan and Depaoli, 2012). In addition, along with the likelihood,  $L(\theta; \mathcal{D})$  which

contains all the relevant *sample-based* information regarding the model parameters, the Bayesian approach also requires probability distribution representing prior beliefs about  $\theta$ , say  $p(\theta)$ .

Combining the likelihood and the prior distribution, Bayes' theorem transforms the prior data beliefs into posterior (or after data) beliefs (Rossi and Allenbi, 2003):

$$p(\theta | \mathcal{D}) \propto L(\theta; \mathcal{D})p(\theta). \quad (9)$$

where  $p(\theta | \mathcal{D})$  is known as the posterior distribution of  $\theta$ , given the data. To be more precise, we have:

$$p(\theta | \mathcal{D}) = \frac{L(\theta; \mathcal{D})p(\theta)}{M(\mathcal{D})}, \quad (10)$$

where

$$M(\mathcal{D}) = \int_{\Theta} L(\theta; \mathcal{D})p(\theta)d\theta, \quad (11)$$

is known as the marginal or integrated likelihood or “evidence” and represents the normalizing constant of the posterior. The marginal likelihood is an important object as it represents the evidence of a given model, in the light of the data, *after* parameter uncertainty has been fully taken into account by integrating the parameter vector out in (11). This represents a key difference to the traditional sampling approach (i.e. frequentist) “in which we consider the data as random and we investigate the behavior of test statistics over imaginary samples from the data generating process that yields  $L(\theta; \mathcal{D})$ ”. The Bayesian would regard the sampling distribution as irrelevant to the problem of inference because “it considers events that have not occurred” (Rossi and Allenbi, 2003, p.305).

The Bayesian approach (as shown in (9)) is also known for its ability to incorporate prior information,  $p(\theta)$ , in the estimation. This is a key advantage of the Bayesian approach, as in addition to the information provided by the data, one can obtain more accurate and reliable parameter estimates by incorporating some “genuine prior information” (Song and Lee, 2012).<sup>2</sup> Within the context of SEM, for instance, a researcher may have information from different sources, such as expert opinion, or result from past studies using similar data, that can be incorporated into the analysis. Such information may range from prior information about the estimates of factor loading from a previous tourism model to the level of correlation between two latent variables (e.g. satisfaction or return intention).

Basically, there are two types of priors: informative and non-informative priors. Informative priors are used when a researcher has good knowledge about the prior distribution from previous studies, while non-informative prior is adopted when we are not in possession of

---

<sup>2</sup> The argument that non-Bayesians do not use prior information is quite wrong. Choosing a model is prior information. Using instrumental variables also involves choices which are equivalent to prior information. Regarding the randomness of  $\theta$ , the purpose of introducing a random variable is because we wish to learn about something unknown. The unknown quantity in statistical studies is the parameter, not the data.

enough prior information to help in drawing posterior inferences. Non-informative priors are also known as “vague” or “diffuse” priors. Some examples of non-informative prior distributions include the uniform distribution over some sensible range of value or the so-called “Jeffrey’s prior” (Kaplan and Depaoli, 2012). Basically, with the use of non-informative priors, Bayesian inference based on the posterior distribution (9) becomes less dependent on the prior distribution,  $p(\theta)$ , and more dependent on the likelihood,  $L(\theta; \mathcal{D})$ . However, even in such case, Bayesian inference is still fundamentally different compared to the frequentist approach, because it is based directly on the posterior in (9) and not on hypothetical “infinite replication of the study (via sampling distributions) that never occurred” (Zyphur and Oswald, 2013, p. 4).

The Bayesian approach has also several other advantages such as performing better in small samples, and providing more accurate statistics for goodness-of-fit and model comparison (Song and Lee, 2012). It can also handle more complicated structural equation models. Before elaborating further on these issues, we provide first some background on Bayesian inference using the Markov Chain Monte Carlo (MCMC) approach.

### 3.1.2. Brief Overview of MCMC Estimation

As the posterior (9) can be highly dimensional, researchers usually summarize information about the posteriors in terms of lower dimensional summary statistics such as the mean and standard deviation. For example, if we have a regression of the form “ $y = \theta_0 + \theta_1 x_1 + \dots + \theta_k x_k$ ”, the posterior mean

$(E[\theta] = \int \theta p(\theta | y, x_1, \dots, x_k) d\theta)$  and the posterior variance of  $\theta_1$  will be used to test hypotheses. A challenge however is that both of these quantities require calculating some multidimensional integrals of the posterior distribution (Rossi and Allenbi, 2003). Historically, the computation of complicated integrations has put the Bayesian approach beyond the reach of many applied researchers (Coelli et al. 2005). Recent developments in powerful simulation algorithms, now provided through several software packages has, however, facilitated the estimation of posterior probability distribution for many models.

One of these most powerful algorithms is MCMC. It is “an iterative process where a prior distribution is specified and posterior values for each parameter are estimated in many iterations” (Zyphur and Oswald, 2013, p. 11). Hence, instead of computing the integrals analytically, one can use simulation-based methods. Specifically, in MCMC approach, we generate a long sample; say  $\{\theta^{(s)}, s = 1, \dots, S\}$  that converges in distribution to the posterior in (11). The normalizing constant  $M(\mathcal{D})$  is not needed as the posterior expectation of an arbitrary vector function of the parameters, say  $g(\theta)$ , can be approximated accurately by:

$$E[g(\theta) | \mathcal{D}] \simeq S^{-1} \sum_{s=1}^S g(\theta^{(s)}). \quad (12)$$

The marginal likelihood,  $M(\mathcal{D})$ , can be approximated as a by-product. There are many ways to do this. One way is to use the Laplace approximation<sup>3</sup>. Since

$$\log M(\mathcal{D}) = \log L(\theta; \mathcal{D}) + p(\theta) - \log p(\theta | \mathcal{D}), \quad (13)$$

Notice that this is an identity in  $\theta \in \Theta$ , where  $\Theta \subset \mathbb{R}^d$  is the parameter space. Therefore, in principle, any specific  $\theta$ , say  $\bar{\theta}$ , may be used to obtain:

$$\log M(\mathcal{D}) = \log L(\bar{\theta}; \mathcal{D}) + \log p(\bar{\theta}) - \log p(\bar{\theta} | \mathcal{D}). \quad (14)$$

Typically, for  $\bar{\theta}$  we can use the posterior mean of  $\theta$ . Both  $\log L(\bar{\theta}; \mathcal{D})$  and  $\log p(\bar{\theta})$  are known and can be easily computed. However,  $\log p(\bar{\theta} | \mathcal{D})$  is unknown. The Laplace approximation assumes that  $p(\theta | \mathcal{D})$  can be approximated by a multivariate normal distribution around the mean and, therefore, we have the following simple expression:

$$\log M(\mathcal{D}) \simeq \log L(\bar{\theta}; \mathcal{D}) + p(\bar{\theta}) + \frac{d}{2} \log(2\pi) + \frac{1}{2} \log |\bar{V}|, \quad (15)$$

where  $\bar{V}$  is the posterior covariance matrix of the parameter vector:

$$\bar{V} = E \left[ (\theta - \bar{\theta})(\theta - \bar{\theta})' | \mathcal{D} \right] \simeq S^{-1} \sum_{s=1}^S (\theta^{(s)} - \bar{\theta})(\theta^{(s)} - \bar{\theta})'. \quad (16)$$

The remaining problem is to implement MCMC, that is to draw a long sample; say  $\{\theta^{(s)}, s = 1, \dots, S\}$ , that converges in distribution to the posterior distribution whose density is in (11). One MCMC technique is the Gibbs sampler. To understand the Gibbs sampler it is, surprisingly, easier to start with the more general family of which the Gibbs sampler is a member; this family is known as the **Metropolis-Hastings algorithm**. The algorithm operates as follows. Suppose we have a starting value, say  $\theta^{(0)}$  and the algorithm is currently at state  $s$  having  $\theta^{(s)}$  as the current MCMC draw. The next draw is produced simply as follows. Given  $\theta^{(s)}$  and  $s = 1, \dots, S$  do:

---

<sup>3</sup> It is, perhaps, useful to mention that well-known model selection criteria such as the AIC and BIC are simply different *asymptotic* approximations to the marginal likelihood.

- Set  $\theta^{(s+1)} = \theta^{(s)} + h\varepsilon$ , if  $\min\left\{1, \frac{p(\theta^{(s+1)} | \mathcal{D})}{p(\theta^{(s)} | \mathcal{D})}\right\} \geq u$ , where  $u$  is a standard uniform random number and  $\varepsilon \sim N_d(0, V)$ , a  $d$ -variate normal distribution.
- Otherwise, set  $\theta^{(s+1)} = \theta^{(s)}$ , and repeat the previous draw.

This is known as the Random Walk Metropolis-Hastings algorithm,  $V$  can be any matrix (for example the covariance matrix from ML or an identity matrix) and  $h$  is a positive constant, which we select by trial-and-error so that approximately  $1/4$  of all draws are finally accepted. The reader familiar with the Simulated Annealing method of maximization will, certainly, notice the similarities.

Another version known as the Independence Metropolis-Hastings algorithm results if we draw a candidate  $\theta^c$  from a (convenient) distribution with density, say  $g(\theta)$ , called the importance, proposal or candidate-generating density. Then we proceed as follows:

- Draw a candidate  $\theta^c$  from a distribution whose density is  $g(\theta)$ .
- Set  $\theta^{(s+1)} = \theta^c$ , if  $\min\left\{1, \frac{p(\theta^c | \mathcal{D}) / g(\theta^c)}{p(\theta^{(s)} | \mathcal{D}) / g(\theta^{(s)})}\right\} \geq u$ , where  $u$  is a standard uniform random number. That is, *accept the candidate*.
- Otherwise, set  $\theta^{(s+1)} = \theta^{(s)}$ , and *repeat the previous draw*.

The **Gibbs sampler** operates by drawing random numbers from the posterior conditional distribution of each parameter given the rest. For example if  $\theta = [\theta_1, \theta_2, \dots, \theta_d]'$ , the Gibbs sampler draws random numbers from the following distributions:

$$\begin{aligned} \theta_1 &| \theta_2, \theta_3, \dots, \theta_d, \mathcal{D}, \\ \theta_2 &| \theta_1, \theta_3, \dots, \theta_d, \mathcal{D}, \\ &(\dots) \\ \theta_d &| \theta_1, \theta_2, \dots, \theta_{d-1}, \mathcal{D}. \end{aligned}$$

Therefore, we have to draw from the posterior conditional distribution with density  $p(\theta_m | \theta_{(-m)}, \mathcal{D})$ ,  $m = 1, \dots, d$ , where  $\theta_{(-m)}$  denotes the parameter vector  $\theta$  with the exception of parameter  $\theta_m$ . It is interesting to note that any element can, in fact, be a vector. If we repeat



this process a large number of times, we obtain a sample  $\{\theta^{(s)}, s = 1, \dots, S\}$  which converges in distribution to (11). However, *the MCMC sample is not i.i.d.*, as we have autocorrelation. This is evident as, for example,  $\theta_1^{(s)}$  and  $\theta_1^{(s-1)}$  cannot be independent. If the autocorrelation is high then, effectively, the Gibbs sampler (or any other MCMC scheme) will not explore the posterior thoroughly in a small number of iterations (say 5,000 or 10,000). Interestingly, the Gibbs sampler does not involve tuning constants like  $h$  or the selection of importance densities so, in this sense, it is automatic and involves only the requirement that one can obtain (easily) random drawings from each posterior conditional distribution.

### 3.1.3. Bayesian Model Comparison

It is common in SEM to compare between different competing models, and to ensure the model fits the data well. The Bayesian approach offers more reliable statistics for goodness-of-fit and model comparison (Lee and Song, 2012). For instance, the model fit indices and model comparison tools (e.g. chi-square, RMSEA, etc.) associated with the covariance-based approach have only asymptotic justification and perform poorly in some complicated modelling conditions. Hence, we expect them to deliver misguided conclusions in small or moderate samples.

We elaborate here on three of Bayesian fit statistics that are very common within the context of Bayesian SEM: the Bayes factor, the Deviance information Criterion and the Posterior predictive p-value. The Bayes factor has been shown to be highly reliable and has many nice statistical properties (Lee and Song, 2012). It has been also extensively adapted within the context of SEMs (Assaf et al. 2016). To introduce the concept of Bayes factor, suppose  $L(\theta; \mathcal{D})$  is the likelihood function of the model where  $\mathcal{D}$  denotes all available data on  $x$  and  $y$ . Denote  $\mathcal{D} = [x_i, y_i; i = 1, \dots, N]$ ,  $D_i \equiv (x_i, y_i) \in \mathbb{R}^{d_p}$ . We assume the data are in deviations about their means to simplify notation. The likelihood function of the SEM is:

$$L(\theta; \mathcal{D}) = (2\pi)^{-\frac{N(p+q)}{2}} |\Sigma(\theta)|^{-N/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^N D_i^* \Sigma(\theta)^{-1} D_i\right\}. \quad (17)$$

The posterior is''  $p(\theta | \mathcal{D}) \propto L(\theta; \mathcal{D}) \cdot p(\theta)$  where  $p(\theta)$  is the prior.

In this context, model comparison becomes easy. If we have two models, say I and II with marginal likelihoods  $M_I(\mathcal{D})$  and  $M_{II}(\mathcal{D})$ , then the Bayes factor in favor of model I and against model II is simply:

$$BF = \frac{M_I(\mathcal{D})}{M_{II}(\mathcal{D})}. \quad (18)$$

If  $BF > 1$  then model I is preferred to model II, in the light of the data. For a number of models, say  $1, 2, \dots, J$  we can obtain marginal likelihoods,  $M_1(\mathcal{D}), M_2(\mathcal{D}), \dots, M_J(\mathcal{D})$ . In turn, we can define posterior model probabilities as follows:

$$\Pi_j \equiv P(\text{model } j \mid \mathcal{D}) = \frac{M_j(\mathcal{D})}{\sum_{j'=1}^J M_{j'}(\mathcal{D})}. \quad (19)$$

Posterior model probabilities can be used for model selection but also for model averaging. For example, if we are interested in parameter  $\theta_1$  and its marginal posterior densities across models are  $p_1(\theta_1 \mid \mathcal{D}), p_2(\theta_1 \mid \mathcal{D}), \dots, p_J(\theta_1 \mid \mathcal{D})$ , the *model-averaged posterior*, which accounts for model uncertainty is:

$$p_*(\theta_1 \mid \mathcal{D}) = \sum_{j=1}^J \Pi_j p_j(\theta_1 \mid \mathcal{D}). \quad (20)$$

Typically, we are interested in Bayes factors but also first and second order posterior moments of the parameter vector  $\theta$  or a vector function  $g(\theta) = [g_1(\theta), \dots, g_M(\theta)]'$ . Generically, the posterior expectation of  $g(\theta)$  is:

$$E[g(\theta) \mid \mathcal{D}] = \int_{\Theta} g(\theta) p(\theta \mid \mathcal{D}) d\theta. \quad (21)$$

Another highly popular Bayesian model comparison tool is the Deviance Information Criterion (DIC), see Spiegelhalter et al. (2002). It is less computationally involved than the Bayes factor and has been used extensively in the field of SEM (Lee and Song, 2012). For example, if we have a competing model  $M_k$ , with a vector of unknown parameter

$$DIC = \overline{D(\theta_k)} + d_k \quad (22)$$

where

$$\overline{D(\theta_k)} = E_{\theta_k} \left\{ -2 \log p(Y \mid \theta_k, M_k) \mid Y \right\} \quad (23)$$

and  $d_k$  here is the effective number of parameters in  $M_k$ . Hence, as shown, the calculation of DIC involves simulating  $\{\theta_k^{(j)}, j = 1, \dots, J\}$  from the posterior distribution. The Winbugs

software we describe can be used to compute DIC. Models with smaller DIC are considered to have a better fit.

Finally, the posterior predictive p-value focuses on the predictive ability of the model in that there “should be little, if any, discrepancy between data generated by the model, and the actual data itself” (Lee and Song, 2014, p.277). To illustrate, assume that  $D(Y | \theta, \Omega)$  is the discrepancy measure between the hypothesized model  $M_o$  and the hypothetical replicate data  $Y^{rep}$ , the posterior predictive p-value is given by:

$$p_B(Y) = \Pr\{Y^{rep} | \theta, \Omega\} \geq D(Y | \theta, \Omega | Y, M_o) \quad (24)$$

A model is considered a good fit if the posterior predictive p-value is close to 0.5. For more details refer to Lee and Song (2012).

#### 4. Monte Carlo Experiment

We provide in Appendix 1 more specific details the Bayesian estimation of a general SEM. Specifically, we provide more details about the priors and posterior distribution and how Bayesian inferences are derived. As the tourism literature lacks such description, we believe this is essential to further describe the Bayesian theorem within the context of SEM<sup>4</sup>.

Before presenting the estimation of a full Bayesian SEM example, we discuss first the results of a Monte Carlo experiment which we conducted to emphasize the power of the Bayesian approach in small samples. As mentioned above, the covariance based approach (i.e. LISREL, AMOS) approach to SEM estimation is only asymptotically true. In other words, it requires large sample to make valid statistical inferences. We conduct here a Monte Carlo simulation to compare between Bayesian and covariance approaches across both small and moderately large sample sizes.

To set up the Monte Carlo experiment: In connection to (1) – (2) suppose we are given actual data on  $x_i$  and  $y_i$  ( $i = 1, \dots, N$ ) and we perform the traditional covariance based approach using maximum likelihood (ML) to find  $\hat{\theta}$ . To proceed with a realistic Monte Carlo experiment, we treat  $\hat{\theta}$  as the true parameter vector and we generate a set of data  $\mathcal{D}^{(r)} = \{x_i^{(r)}, y_i^{(r)}, i = 1, \dots, N\}$  for replications  $r = 1, \dots, R$ . To generate  $\eta_i$  we use the reduced form:

$$\eta_i = (I - B)^{-1} (\Gamma \xi_i + \zeta_i), \quad i = 1, \dots, N.$$

---

<sup>4</sup> To reduce the amount of technical details we have moved the Bayesian description of the linear SEM to Appendix 1

The generation of  $\xi_i$  and  $\zeta_i$  is straightforward. Given  $\eta_i$  we can easily generate  $y_i$  and  $x_i$  for each replication of the Monte Carlo experiment.

For each generated data set we perform again ML and we also perform Bayesian analysis using the MCMC algorithm in Appendix 2<sup>5</sup>. For the MCMC algorithm the number of draws is set to  $S = 6,000$  of which we discard the first 1,000 to mitigate the impact of start-up effects. Our starting value is always the ML estimator  $\hat{\theta}$ . Whenever Geweke's (1993) convergence diagnostic indicates non-convergence, we take another 2,000 iterations and look again at Geweke's statistic.

We use flat priors on all parameters, we assume all covariance matrices are diagonal, and we repeat the Monte Carlo experiment for  $R = 10,000$  replications. In the 10,000 replications we found non-convergence in 322. In all of them taking another 5,000 iterations was sufficient. In the vast majority, however, 1,000 – 2,000 additional draws were found enough. We implement ML using a standard Gauss-Newton algorithm with analytic gradient and Hessian, which is also of use in the GC – MCMC algorithm. For ML we set the maximum number of iterations to 500; if the limit is exceeded we generate another data set to perform ML but Bayesian MCMC analysis is performed anyway with the data set where ML failed to converge. We believe this gives to ML a fair advantage. The number of iterations was exceeded in 812 cases out of the 10,000.

To perform the Monte Carlo experiment we relied on a well-established model on lodging brand equity (Figure 1), previously published in Hsu et al. (2012). The model and items used for measurement are well discussed in their paper, so we do not intend to reiterate them here. Based on Figure 1, the  $\xi$  s in our case are perceived quality, brand awareness, brand image, management trust, and brand reliability. The  $\eta$  s are brand loyalty and brand choice intention. For  $\xi$  s we have 16 indicators. For example loyalty is measured through three indicators (BL1, BL2, BL3), intention through three indicators (BR1, BR2 and BR3), etc<sup>6</sup>. All observed variables are on a Likert scale (1-7). To generate the data for a specific replication we use the following strategy:

- a) We estimate the model by maximum likelihood (ML) assuming all covariance matrices are diagonal.
- b) Using the estimated parameters we generate  $\eta_i, \xi_i, x_i, y_i$  as described above.

---

<sup>5</sup> As the Gibbs sampler is, typically, hard to converge to the posterior if the data are highly correlated or under anomalies are at work, an alternative is to use techniques that utilize first- and second-order derivative information from the log posterior. The algorithm we use is not unlike the Random Walk Metropolis-Hastings algorithm but it performs much better as it uses first and second derivative information from the posterior. It is the Girolami and Calderhead (2011, GC) algorithm, see Appendix B.

<sup>6</sup> For more details about these indicators refer to Hsu et al. (2012).

- c) Since the data is continuous we transform to Likert scale using the minimum and maximum values of the continuous data. The covariance matrix is recomputed using the new ordinal data.

The results for both the structural and measurement models across different sample sizes are presented in Table 1. We tried both small sample sizes (N=75, 150) as well as moderately large sample sizes (N=200, and 300) which we consider typical in empirical studies. In each case, we show the root mean square error (RMSE) of each parameter estimate for both ML and Bayesian approaches, where a smaller RMSE indicates a better performance.

The results clearly indicate that there is a significant gain from the Bayesian approach across all sample sizes. For instance, we do not observe any single instance where the Bayesian approach generate larger RMSE. This comes to support previous findings from the literature that the Bayesian approach outperforms the traditional covariance based approach, particularly for small sample size (e.g. Lee and Song, 2004). We also believe that such finding is critically important for the tourism literature as it would eliminate the need to continuously collect large samples of data.

## **5. Bayesian Estimation of SEM: A Model of Social Exchange Theory (SET)**

As we discussed before, Bayesian inference in SEM requires, first, deriving the conditional posteriors, and then setting up the MCMC procedure (as explained in 3.1.2) to simulate from the conditional posteriors and obtain statistical inferences. This can be still highly challenging for the applied researcher and requires some heavy computer coding. Fortunately, now certain SEM software packages provide Bayesian inference in SEM. However, these can be highly inflexible in terms of adjusting the prior distribution of the SEM parameters, or in terms of estimating more advanced version of SEMs. We encourage tourism researchers to use the Winbugs software, which is very useful for a wide range of statistical models including SEM. The advantage of the Winbugs software is that it helps the researcher “really concentrate on building and refining an appropriate model without having to invest large amounts of time in coding up the MCMC analysis and the associated processing of the results” (Griffin and Steel, 2005, p.164). The algorithm in Winbugs has been mainly developed using MCMC, and the software necessitates only coding the model and the prior so it requires a much smaller investment on part of the user.

We do not intend here to provide a detailed description of the Winbugs software as this has been provided in several textbooks on the topic (Ntzoufras, 2008), but we describe here its main outputs as part of our application. We also provide the Winbugs code we used to estimate the model in order to guide future tourism applications using the software. Winbugs provides some useful convergence diagnostics, as well as some model comparison tools such as DIC.

Our illustrative application is a SET model published by Jeong and Oh (in press) who used the model to examine the prevalent business-to-business (B2B) relationship between destination management companies (DMCs) and meeting planners (MPs). While Jeong and Oh provide an

extensive review and background on both the illustrative model and the DMC-MP B2B relationship, we briefly recapitulate them here for the purpose of introducing our illustration. In general, DMCs and MPs work closely to attract various event and meeting businesses to target destinations (Sautter & Leisen, 1999). DMCs are typically destination-bound and serve MPs with local knowledge and resources needed to execute events, while MPs bring to DMCs an extensive market coverage beyond the DMC's location. These two business entities have often formed both formal and informal partnerships over a long period of time, which may afford both partners an opportunity to build mutual dependence and trust and, hence, qualify an exemplary setting for SET applications.

Following a thorough review of key variables of SET by Lambe, Wittman, and Spekman (2001), Jeong and Oh (in press) proposes a SET model to examine the B2B relationship between DMCs and MPs (see Figure 2). For the purpose of our illustration in this paper, however, we reanalysed the same model from the perspective of MPs in particular. The model closely follows Morgan and Hunt's (1994) trust-commitment framework that has been widely used to explain B2B relationships. Jeong and Oh's proposed model additionally included the concept of relationship satisfaction as another key mediating variable to enrich the model's explanatory power. This SET model aims to predict the relationship partners' long-term as well as short-term commitment to the focal relationship. Thus, both trust and relationship satisfaction mediate the effects of the four independent latent variables (communication quality, opportunistic behavior, financial dependence, and social dependence) on relationship commitment and propensity to leave the relationship (see Anderson & Narus, 1990; Claycomb & Franwick, 1997; Gundlach et al., 1995; Morgan & Hunt, 1994; Nevin, 1995). For additional background including the conceptual definitions, variable operationalizations, and the theoretical relationships in the model, refer to Jeong and Oh (in press).

All model variables were operationalized as latent variables measured with the multiple items that were extracted from previous studies and a series of preliminary studies (Jeong & Oh, in press). Except for propensity to leave and relationship commitment, all the other variables were anchored on a 5-point Likert scale. Three items, operationalized each on a 5-point 'very dissatisfactory-very satisfactory,' 'terrible-delightful,' and 'of low/high value' scale, measured the partner's overall satisfaction with the current DMC-MP business relationship. Propensity to leave was a three-item battery measuring the partner's intention to leave the current relationship in the next six months, one year, or two years, on a very low-very high likelihood scale. Table 2 summarizes the measurement items and Jeong and Oh provides more detail. After deleting all missing values, the final sample included 101 observations.

Before presenting the Bayesian results, we note that we attempted to estimate the model first using the traditional covariance based approach with Mplus. However, the model did not converge due, most likely, to the small sample size (or more precisely, a small sample relative to the number of parameters). We show below that MCMC converged well with this model and resulted as well in a strong model fit. This comes to further support the results from our simulation that the Bayesian approach performs better than the traditional approach in small sample sizes. The correlation matrix between all latent variables and the Bayesian results for the measurement model are presented in Tables 2 and 3, respectively. Before discussing the results,

we first checked the convergence of MCMC chains using Winbugs (Figure 3). For example, as shown in the convergence plots of some model parameters the chains have mixed well after few thousands iteration in each case.

With the Bayesian approach, we report results in terms of the posterior distribution. For instance, the posterior mean and the posterior SD are presented in Table 2. The loadings were all statistically significant at the 5% level, as noted by the low standard deviation<sup>7</sup>. The results from the structural model are presented in Table 4. For each relationship, we show the posterior mean and standard deviation, as well as 90% and 95% higher posterior densities. Figure 4 also presents the plots of the empirical posterior distributions for some these relationships.

As shown, except the impact of communication quality on relationship satisfaction, all other relationships are significant at either the 5% or the 10% level. The results seem to be also theoretically sound. Communication quality had a significant, positive relationship with trust whereas opportunistic behaviour was negatively related to both trust and relationship satisfaction. As expected, a significant, positive relationship existed between financial dependence and trust. Although the effect of social dependence on trust was “insignificant”,<sup>8</sup> its effect on relationship satisfaction was significant and positive supporting the research hypothesis of interest. Trust was a significant, negative antecedent of propensity to leave but a positive determinant of relationship commitment. Finally, relationship satisfaction had a significant negative association with propensity to leave and a significant positive association with relationship commitment.

To ensure the validity of our hypothesis tests, and to confirm the Bayesian model is performing well with this small sample size, we also assessed the overall fit of the model using the posterior predictive p-value. For example, we found that the posterior predictive p-value is 0.58 which confirms that the model fits the data well. We also compared the model in Figure 2 against another competing model, which allows also for direct relationships between communication quality, opportunistic behaviour, social dependence, financial dependence and propensity to leave and relationship commitment respectively.

---

<sup>7</sup> With Bayesian, it is more appropriate to look at the prediction intervals to assess significance. We confirmed that all these loadings are “significant” in the sense of footnote 8.

<sup>8</sup> “Insignificant” in the Bayesian paradigm means that the so-called 95% highest-posterior-density-interval (HPDI) does *not* include zero. We use the term for brevity as there is no such thing as “statistical significance” in the Bayesian paradigm. Moreover, “significant” in the Bayesian paradigm means that the 95% HPDI does not include zero.

Using the DIC and Bayes factor (see section 3.1.3) we showed that the model in Figure 2 generally outperforms this competing model. For instance, the Bayes factor in favor of our model was 8.12 (see equation 18), indicating a significantly better performance than the competing model<sup>9</sup>.

## 6. Other Potential Extensions SEM using the Bayesian Approach

So far we have only discussed examples of linear SEM applications. In this section, we briefly outline other extensions of SEM where the Bayesian approach has proven to be highly powerful. We believe it is important to shed light on these models to encourage more advanced SEM applications in tourism research. Unfortunately, the heavy reliance on the covariance based SEM approach creates limitations in estimating some of these models.

### 6.1. Finite Mixture SEM

A well-known estimation problem that has been ignored in tourism research is the issue of unobserved heterogeneity (Assaf et al. 2016). Assuming that the data are always homogenous may more often than not lead to biased and wrong conclusions. The problem with assuming homogeneity can be illustrated by taking an example on customer satisfaction. Consider the relationship between customer satisfaction and causes of customer satisfaction (e.g. service effectiveness, service quality, and promotion). Suppose that in the data there are three distinct groups of customers. What distinguishes these groups is that their level of customer satisfaction is determined differently. Some customers may find that service effectiveness is the most important (labelled as group  $g = 1$ ), whereas other customers ( $g = 2$ ) may find that service quality is the key to their satisfaction level. The satisfaction of the last group of customers ( $g = 3$ ) is mostly determined by promotion. Hence, failing to understand this heterogeneity would lead researchers to analyse these data as if they were homogeneous. The goodness-of-fit indices would not reveal that the model was incorrectly specified and the researcher would not be alerted to the unaccounted heterogeneity in the model. Furthermore, the structural parameter estimates would be seriously biased. In other words, not accounting for unobserved heterogeneity has the same implications as misspecification in regression analysis. In case heterogeneity exists an important extension to the linear SEM is the finite mixture SEM which can be written as:

$$\eta_i | \xi_i, I_i = g \sim N_m \left( \Pi_g \xi_i, \Omega_{\eta, g}(\theta) \right), \quad (25)$$

---

<sup>9</sup> As indicated in (18), if Bayes factor  $> 1$  then model I is preferred to model II.



where  $\Omega_{\eta,g}(\theta) = (\mathbf{I} - \mathbf{B}_g)^{-1} \Psi (\mathbf{I} - \mathbf{B}_g)^{-1'}$ , and  $I_i$  represents a discrete random variable taking values in  $\{1, \dots, G\}$ , with probabilities  $P(I_i = g | \xi_i) = \pi_g$ ,  $g = 1, \dots, G$ ,  $\pi_g \geq 0$ ,  $\sum_{g=1}^G \pi_g = 1$ . Here,  $g$  denotes the particular group and can take values  $1, 2, \dots, G$  where  $G$  denotes the number of groups. We refer the reader to Assaf et al (2016) and Lee and Song (2012) for more details about this model. While the finite mixture model can be estimated using the covariance-based (i.e. traditional) approach, the Bayesian approach is better suited to correctly identify the number of groups in the data (Richardson and Green, 1997). As highlighted by Lee and Song (2012), using the traditional estimation method in estimating finite mixture SEMs can be problematic in terms of identifying the number of groups due to non-regularity problems<sup>10</sup>.

## 6.2. Non-parametric and Semi-Parametric SEMs

Both non-parametric and semi-parametric SEMs have also been heavily ignored in the tourism literature, despite being more appropriate in handling non-normal data. The fact that traditional SEMs also assume that the latent variables follow a normal distribution can be also be problematic. As the latent variables are unobserved, it is impossible to check whether this is a valid assumption. One way to relax this assumption, and avoid having spurious statistical results is to use the semi-parametric or non-parametric SEM, where again, the Bayesian approach has been shown to be highly powerful. For some detailed studies on the topic refer to Lee and Song (2008), and Song et al. (2009).

## 6.3. SEMs with Continuous and Ordered Categorical Variables

The Bayesian approach also offers high flexibility in handling models with continuous and ordered categorical variables. Most SEM applications in tourism are often based on the use of Likert scale data, where satisfying normality can be an issue. For instance, to claim normality of these Likert scale data we need most answers to be in the middle category. However, in some cases, this requirement is not satisfied.

The common approach is to treat all observed variables as continuous data coming from a normal distribution. However, this can lead to spurious results if the distribution of these observed variables does not follow, approximately, a normal distribution (for example, when most respondents select categories at both ends). An arguably better way to analyze such type of data is to treat them as observations that come from a latent continuous distribution with a threshold specification. So far, we have never seen such approach adopted in the tourism literature. For example, if we have left skewed data, the threshold approach for analyzing such

---

<sup>10</sup> The Bayesian finite mixture model can also be estimated using the Winbugs software (see Assaf et al. 2016 for coding details).

type of data is “to treat the ordered categorical data as manifestations of an underlying normal variable  $y$ ”(Song and Lee, 2012, p.87).

$$z = m, \quad \text{if } \alpha_m \leq y \leq \alpha_{m+1} \quad (26)$$

where  $\alpha$ 's are the thresholds,  $z$  is the observed ordered categorical variable, and  $m$ 's represent the observed values for  $z$ .

Analysing such a model is not trivial and involves computing multiple complicated integrals. A multistage method using generalized least square (GLS) has been proposed in the literature to analyse (26). However, other studies (Shi and Lee, 2000) have discussed the problem of reaching an optimal solution with such approach. With the Bayesian approach one can handle more effectively (26). Using the idea of data augmentation in MCMC one can simply augment the observed data with the latent continuous measurement corresponding to these ordered categorical variables in the posterior analysis. In other words, one can treat the underlying continuous measurement as missing data or parameters, and then one can augment them with the observed data in the posterior analysis. Hence, the model that is based on the complete dataset becomes one with continuous variable. For more discussion on the topic, refer to Dunson (2000), Lee and Song (2014) and Lee and Song (2012).

#### 6.4. Transformation SEMs

When the data is highly non-normal, even non-parametric and semi-parametric SEM can face some challenges (Lee and Song, 2012). As indicated above, satisfying normality is the one the main assumptions of SEMs. Fortunately, some transformation models have been developed in a Bayesian framework to address highly skewed data. The idea is to use a transformation SEM defined by:

$$f(y_i) = \mu + \Lambda\omega_i + \varepsilon_i \quad (27)$$

where  $f_j(\cdot)$  is a transformation function that can be used to generate a normal distribution or to address extreme skewness so that the resulting model meets the normality assumption in SEM. With the Bayesian approach  $f_j(\cdot)$  can be approximated using Bayesian P-splines (see Lee and

Song, 2014 and Lee and Song, 2012 for more details)<sup>11</sup>. In some cases even Box-Cox transformations may suffice.

## 7. Concluding Remarks

The aim of this paper was to provide a comprehensive introduction of the Bayesian approach for SEM estimation. Despite receiving a strong attention across other related fields, the use of the Bayesian approach is still highly limited in the tourism literature. We highlighted in this paper the power of the Bayesian approach and discussed its distinctive difference from the traditional covariance-based approach to SEM estimation.

Overall, we believe there are five main reasons why tourism researchers might select the Bayesian approach for SEM estimation. First, some complicated models such as the ones discussed in the previous section are harder to converge with traditional methods (e.g. mixture models; non-normal models, etc.), and some models are not even possible to estimate. Bayesian statistics can also help in model identification and result in more accurate parameter estimates (Depaoli, 2013; 2014).<sup>12</sup> Second, “many scholars prefer Bayesian statistics because they believe population parameters should be viewed as random” (Depaoli and van de Schoot, 2015, p.3). Third, with the Bayesian approach one can prior information into the estimation. Fourth, as highlighted several times above, the Bayesian statistics is not based on large samples and hence may generate reasonable results even with small to moderate sample sizes. This was also reinforced by the results of our Monte Carlo simulation. Fifth, and finally, the Bayesian approach offers more accurate and less sensitive fit statistics and model comparison tools.

Despite all these advantages, the main goal should not be understood as encouraging some naïve applications of the Bayesian approach, or even using the Bayesian approach in the interest of “mathematistry”. We understand that most researchers in tourism are usually more comfortable using the frequentist approach for SEM estimation. As indicated by Depaoli and van de Schoot (2015), using the Bayesian approach without good knowledge of the method can be dangerous,

---

<sup>11</sup> In addition to the models discussed in this section, we note that the Bayesian approach can also handle effectively latent curve and longitudinal data. Traditionally, most traditional longitudinal SEMs have focused on univariate observed variables measured over repeated periods of time. The field has been slow in developing more advanced longitudinal models, probably because the covariance matrix of observed and latent variance involved in different time periods can be complicated (Lee and Song, 2012). Recently some more advanced Bayesian models have been developed for analyzing longitudinal data, which relaxes the univariate assumption. This is an interesting topic for future research.

<sup>12</sup> One simple example is linear regression  $y=X\beta+u$ , when  $X$  is collinear or even singular. Then the matrix  $X'X$  cannot be inverted or, if it can be inverted, standard errors will be very large. A simple normal prior yields the estimator  $b=(X'X+gI)^{-1}X'y$  where  $g$  is related to prior information. In the frequentist approach this is known as “ridge regression”: One mechanically adds a small constant,  $g$ , to the cross-products matrix to make it better behaved. However, there is a clear Bayesian interpretation of this mechanical procedure.

particularly in terms of interpreting the Bayesian features and / or results. The Bayesian approach can also be sensitive to the selection of appropriate priors –but this is an empirical matter. From here, conducting sensitivity analysis to check whether the results are stable across prior choices becomes essential (Assaf et al. 2016). There are also other important steps that should be checked when using the Bayesian approach- we refer the reader to the study of Depaoli and van de Schoot (2015) for more details.

## References

- Anderson, J.C., & Narus, J.A. (1990). A model of distributor firm and manufacturer firm working partnerships. *Journal of Marketing*, 54(January), 42-58.
- Anderson, J.C., & Weitz, B. (1989). Determinants of continuity in conventional industrial channel dyads. *Marketing Science*, 8(4), 310-323.
- Ariani, D. W. (2015). Relationship model of personality, communication, student engagement, and learning satisfaction. *Business, Management and Education*, 13(2), 175-202.
- Assaf, A. G., Oh, H., & Tsionas, M. G. (2016). Unobserved heterogeneity in hospitality and tourism research. *Journal of Travel Research*, 55(6), 774-788.
- Ball D., Coelho, P.S., & Machás, A. (2004). The role of communication and trust in explaining customer loyalty: An extension to the ECSI model. *European Journal of Marketing*, 38(9/10), 1272-1293.
- Bansal, H. S., Irving, P. G., & Taylor, S. F. (2004). A three-component model of commitment to service providers. *Journal of the Academy of Marketing Science*, 32(3), 234–50.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables* (JohnWiley & Sons, New York). NY, USA.
- Blau, P.M. (1968). Interaction: Social exchange. *International Encyclopedia of the Social Sciences*, 7, 452-458.
- Claycomb, V., & Franwick, G.L. (1997). The dynamics of buyers' perceived costs during the relationship development process. *Journal of Business-to-Business Marketing*, 4(1), 1-37.
- Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.
- Cropanzano, R., Anthony, E., Daniels, S., & Hall, A. (2016). Social exchange theory: A critical review with theoretical remedies. *Academy of Management Annals*, 11(1), 479-516.

- Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods*, 18, 186-219.
- 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
- Depaoli, S., & van de Schoot, R. (2015). Improving Transparency and Replication in Bayesian Statistics: The WAMBS-Checklist.
- Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B*, 62, 355 –366.
- Dwyer, F.R., Schurr, P.H., & Oh, S. (1987). Developing buyer-seller relationships. *Journal of Marketing*, 51(April), 11-27.
- Frazier, G.L. (1983). Interorganizational exchange behavior in marketing channels: A broadened perspective. *Journal of Marketing*, 47(Fall), 68-78.
- Ganesan, S. (1994). Determinants of long-term orientation in buyer-seller relationships. *The Journal of Marketing*, 1-19.
- Geysken, I., Steenkamp, J.E., & Kumar, N. (1999). A meta-analysis of satisfaction in marketing channel relationships. *Journal of Marketing Research*, 36(May), 223-238.
- Geweke, J. (1993). Bayesian treatment of the independent student-t linear model. *Journal of applied econometrics*, 8(S1).
- Girolami, Mark, and Ben Calderhead. "Riemann manifold langevin and hamiltonian monte carlo methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.2 (2011): 123-214.
- Griffin, J. E., & Steel, M. F. (2007). Bayesian stochastic frontier analysis using WinBUGS. *Journal of Productivity Analysis*, 27(3), 163-176.
- Gundlach, G.T., Achrol, R.S., & Mentzer, J.T. (1995). The structure of commitment in exchange. *Journal of Marketing*, 59(January), 35-46.
- Homans, G. C. (1958). Social behavior as exchange. *American journal of sociology*, 63(6), 597-606.
- Hsu, C. H., Oh, H., & Assaf, A. G. (2012). A customer-based brand equity model for upscale hotels. *Journal of Travel Research*, 51(1), 81-93.
- Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. *Handbook of structural equation modeling*, 650-673
- Lambe, C. J., Wittmann, C. M., & Spekman, R. E. (2001). Social exchange theory and research on business-to-business relational exchange. *Journal of Business-to-Business Marketing*, 8(3), 1-36.

- Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, 39(4), 653-686.
- Lee, S. Y., Lu, B., & Song, X. Y. (2008). Semiparametric Bayesian analysis of structural equation models with fixed covariates. *Statistics in Medicine*, 27(13), 2341-2360.
- Lee, S. Y., & Song, X. Y. (2014). Bayesian structural equation model. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(4), 276-287.
- Lee, S. Y., & Song, X. Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences*. John Wiley & Sons.
- Liker, J. K., & Choi, T. Y. (2004). Building deep supplier relationships. *Harvard business review*, 82(12), 104-113.
- Mohr, J., & Nevin, J.R. (1990). Communication strategies in marketing channels: A theoretical perspective. *Journal of Marketing*, 54(4), 36-51.
- Moorman, C., Deshpande, R., & Zaltman, G. (1993). Factors affecting trust in market research relationships. *Journal of Marketing*, 57(January), 81-101.
- Morgan, R.M., & Hunt, S.D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, 58(July), 20-38.
- Nyaga, G.N., Whipple, J.M., & Lynch, D.F. (2010). Examining supply chain relationships: Do buyer and supplier perspectives on collaborative relationships differ? *Journal of Operations Management*, 28, 101-114.
- Nevin, J.R. (1995). Relationship marketing and distribution channels: Exploring fundamental issues. *Journal of the Academy of Marketing Science*, 23(4), 327-334.
- Moorman, C., Deshpande, R., & Zaltman, G. (1993). Factors affecting trust in market research relationships. *The Journal of Marketing*, 81-101.
- Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS* (Vol. 698). John Wiley & Sons.
- Reisinger, Y., & Turner, L. (1999). Structural equation modeling with Lisrel: application in tourism. *Tourism Management*, 20(1), 71-88.
- Richardson S. and Green P.J. 1997. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society B* 59: 731-792.
- Rossi, P. E., & Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science*, 22(3), 304-328.
- Sautter, E.T. & Leisen, B. (1999). Managing stakeholders a tourism planning model. *Annals of Tourism Research*, 26(2), 312-328.

Shi, J. Q., & Lee, S. Y. (2000). Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1), 77-87.

Song, X. Y., Xia, Y. M., & Lee, S. Y. (2009). Bayesian semiparametric analysis of structural equation models with mixed continuous and unordered categorical variables. *Statistics in medicine*, 28(17), 2253-2276.

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.

Zyphur, M. J., & Oswald, F. L. (2013). Bayesian probability and statistics in management research: A new horizon. *Journal of Management*, 39(1), 5-13.

Figure 1. A Customer-Based Lodging Brand Equity Model

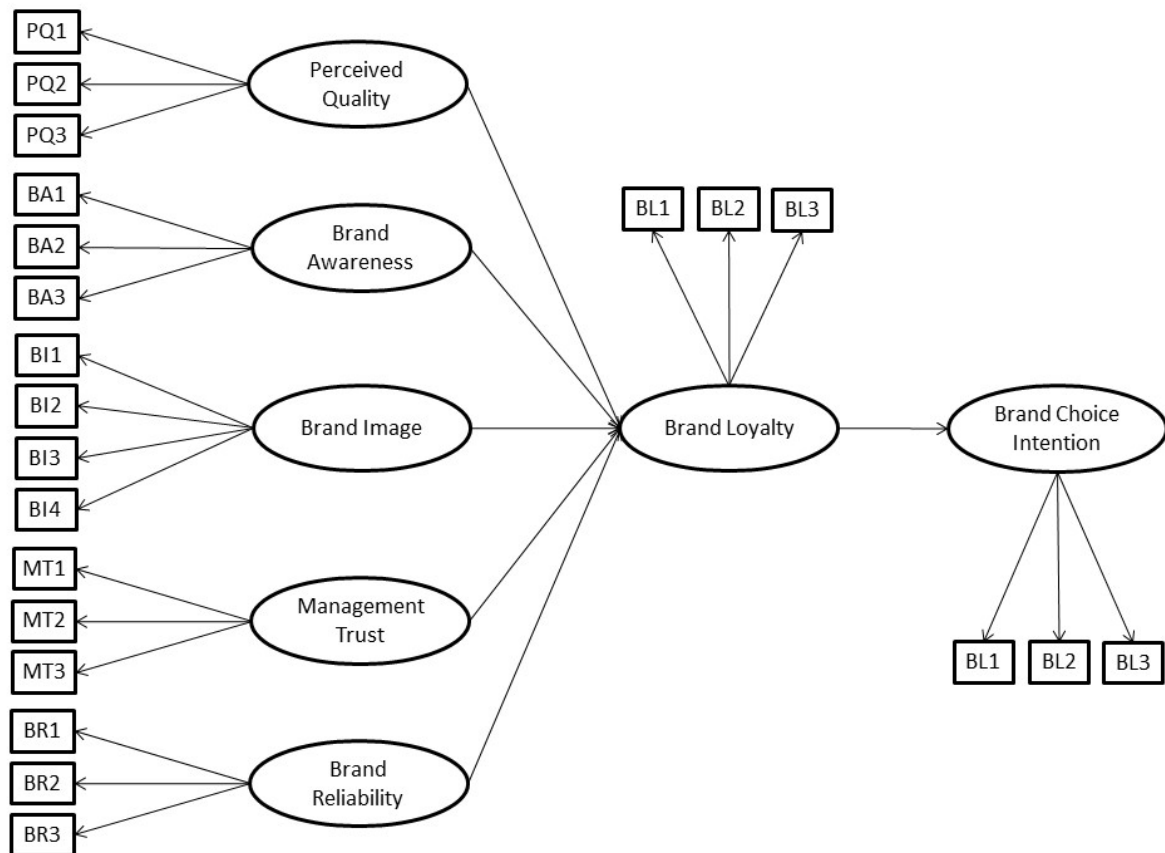


Table 1. Monte Carlo results, RMSE of parameters

	N=75		N=150		N=200		N=300	
parameter	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes
From perceived quality								
PQ1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PQ2	0.321	0.158	0.225	0.114	0.173	0.093	0.155	0.087
PQ3	0.244	0.141	0.189	0.112	0.145	0.095	0.132	0.076
From brand awareness								
BA1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BA2	0.222	0.132	0.187	0.082	0.144	0.077	0.129	0.054
BA3	0.250	0.137	0.210	0.097	0.152	0.081	0.133	0.071
From brand image								
BI1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BI2	0.278	0.112	0.213	0.115	0.157	0.091	0.132	0.083
BI3	0.302	0.130	0.225	0.109	0.176	0.085	0.145	0.074
BI4	0.244	0.132	0.212	0.125	0.194	0.098	0.173	0.095
From management trust								
MT1	0.000	0.00	0.000	0.000	0.000	0.000	0.000	0.000
MT2	0.344	0.141	0.188	0.081	0.133	0.065	0.101	0.055
MT3	0.289	0.137	0.203	0.087	0.136	0.071	0.100	0.062
From brand reliability								
BR1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BR2	0.303	0.144	0.271	0.188	0.210	0.103	0.173	0.087
BR3	0.278	0.115	0.213	0.175	0.184	0.102	0.172	0.088
From brand loyalty								
BL1	0.317	0.286	0.283	0.214	0.210	0.165	0.188	0.111



BL2	0.385	0.277	0.317	0.220	0.265	0.171	0.210	0.133
BL3	0.322	0.220	0.288	0.201	0.215	0.177	0.178	0.130
From brand choice intention								
BC1	0.310	0.244	0.282	0.211	0.265	0.189	0.231	0.164
BC2	0.303	0.271	0.285	0.212	0.271	0.187	0.214	0.171
BC3	0.289	0.214	0.277	0.210	0.273	0.188	0.221	0.168
To Brand Loyalty								
Perceived Quality	0.442	0.228	0.371	0.187	0.315	0.122	0.280	0.084
Brand Awareness	0.389	0.187	0.313	0.144	0.288	0.115	0.265	0.087
Brand Image	0.335	0.165	0.285	0.132	0.211	0.105	0.189	0.085
Management Trust	0.401	0.213	0.387	0.154	0.222	0.115	0.277	0.086
Brand Reliability	0.423	0.357	0.388	0.132	0.285	0.106	0.255	0.087
To brand choice intention								
Brand Loyalty	0.515	0.314	0.473	0.289	0.412	0.233	0.380	0.215
variance parameters <sup>(a)</sup>	0.345	0.285	0.312	0.217	0.296	0.180	0.275	0.164

**Notes:** For Bayes MCMC analysis we use flat priors on all coefficients. Zero entries correspond to coefficients normalized to unity. (a) Reported is the average RMSE of all variance parameters in the SEM.

Figure 2. Proposed Model of Social Exchange Theory and an Extension

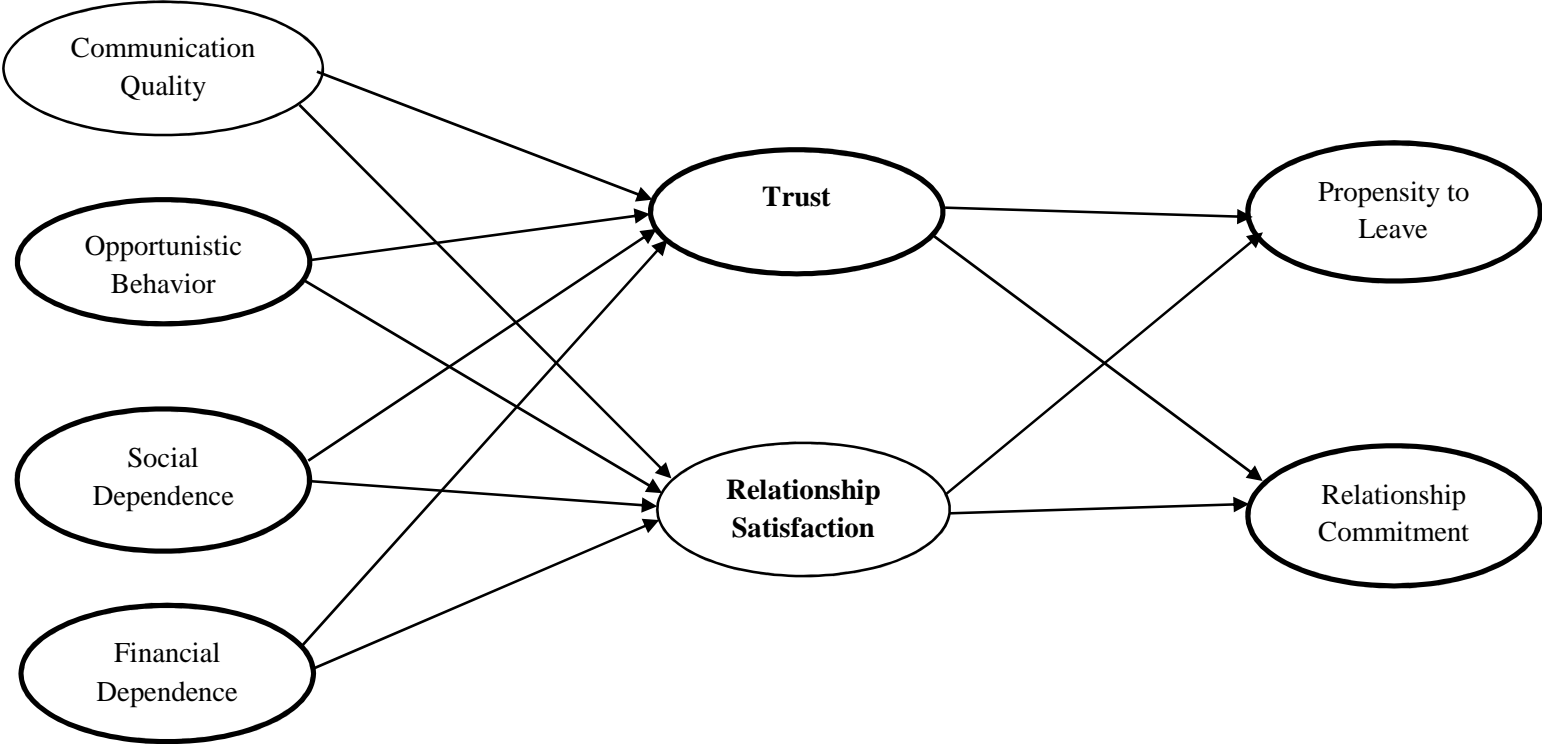


Table 2. Measurement Model Results

Measures (abbreviated)	Posterior Mean	Posterior S.D
Communication Quality		
This major BUSINESS partner ... communicates well their expectations about our firm performance	1.000	
... frequently discusses with us the business ideas that can benefit mutually	1.277	0.162
... is good at notifying us about potential business opportunities	1.407	0.175
... is helpful in providing feedback on our performance	1.254	0.162
Opportunistic Behavior		
Sometimes this major BUSINESS partner... promises to do things without actually doing them later.	1.000	
... gets information from us and contacts our vendors directly later	1.040	0.156
... works with my BUSINESS and our competitors simultaneously to maximize own benefits	1.203	0.187
... tends to treat my BUSINESS as one tentative option while considering other BUSINESS as alternatives	1.295	0.189
Financial Dependence		
The relationship with this major BUSINESS partner ... is built upon frequent business transactions	1.000	
... is motivated mainly by collaborative business opportunities		
... is based on mutual financial gains	1.036	0.226
Social Dependence		
... is based largely on a shared feeling of being “on the same boat” for our respective businesses	1.000	
... embraces our close friendship in its center		
... is built rather on our personal networking and acquaintance	1.118	0.263
Trust		
In our relationship, this major BUSINESS partner ... can be trusted	1.000	
... can be counted on to do what is right	0.903	0.102
... has high integrity	0.954	0.102
... is a very reliable business partner	0.979	0.114
... is consistent in the manner the partner conducts the business with my BUSINESS	0.978	0.110
Relationship Satisfaction		
The overall relationship with this major BUSINESS partner has been ... very dissatisfactory – very satisfactory	1.000	
... terrible – delightful	0.951	0.129
... of no value – of very high value	0.921	0.122
Propensity to Leave		
The chances of terminating the relationship with this major BUSINESS partner ... within the next six months?	1.000	
... within the next one year?	1.159	0.105

... within the next two years?	0.863	0.105
<b>Relationship Commitment</b>		
The relationship with this major BUSINESS partner ... is something we are very committed to	1.000	
... is something my BUSINESS intends to develop more in the future	1.022	0.103
... deserves my BUSINESS' maximum effort to maintain	0.959	0.138
... is something that my BUSINESS will continue devoting necessary resources to strengthen	0.932	0.102

*Note:* All items were measured on a 5-point scale; the relationship satisfaction items were anchored on the three scale labels directly, the propensity to leave items on a very low - very high scale, and all the other construct items on a strongly disagree – strongly agree Likert scale.

Table 3. Correlation Matrix

MP Group	$\alpha$	$\rho_{\eta}$	AVE	1	2	3	4	5	6	7	8
1. Communication Quality	0.91	0.94	0.78	1.00							
2. Opportunistic Behavior	0.88	0.92	0.74	-0.42	1.00						
3. Financial Dependence	0.73	0.88	0.79	0.38	-0.14	1.00					
4. Social Dependence	0.71	0.87	0.77	0.27	-0.17	0.28	1.00				
5. Trust	0.95	0.96	0.83	0.52	-0.47	0.41	0.30	1.00			
6. Relationship Satisfaction	0.87	0.92	0.79	0.49	-0.43	0.30	0.42	0.66	1.00		
7. Propensity to Leave	0.91	0.94	0.85	-0.23	0.57	-0.26	-0.05	-0.54	-0.50	1.00	
8. Relationship Commitment	0.88	0.92	0.74	0.45	-0.31	0.40	0.40	0.65	0.61	-0.36	1.00

$\alpha$  = Cronbach's alpha of internal consistency;  $\rho_{\eta}$  = composite reliability; AVE = amount of variance extracted.

Figure 3. MCMC Convergence for Some Model Parameters

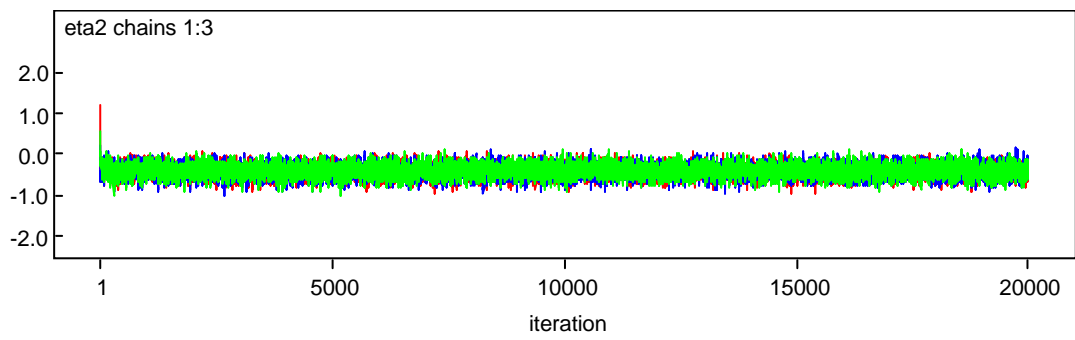
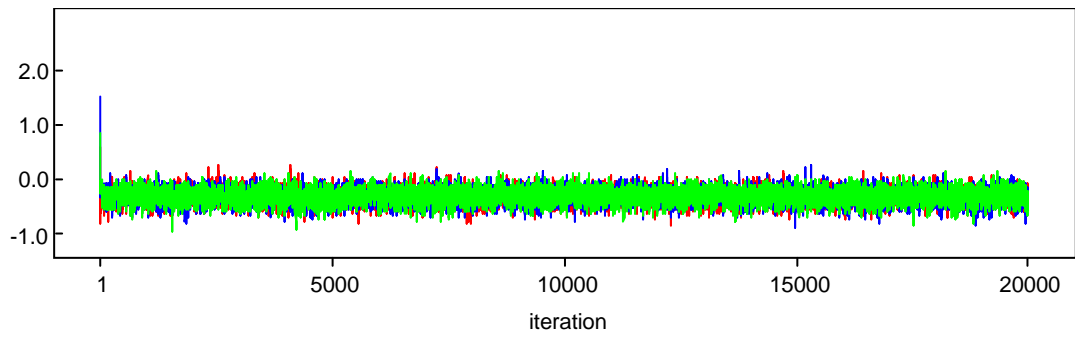
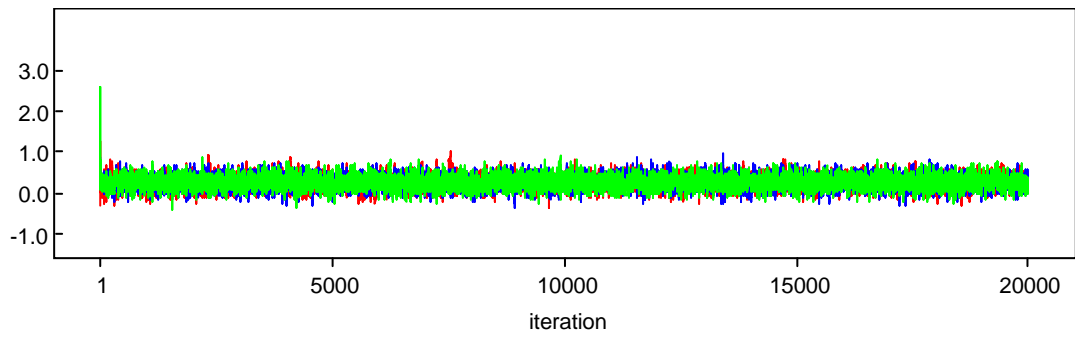
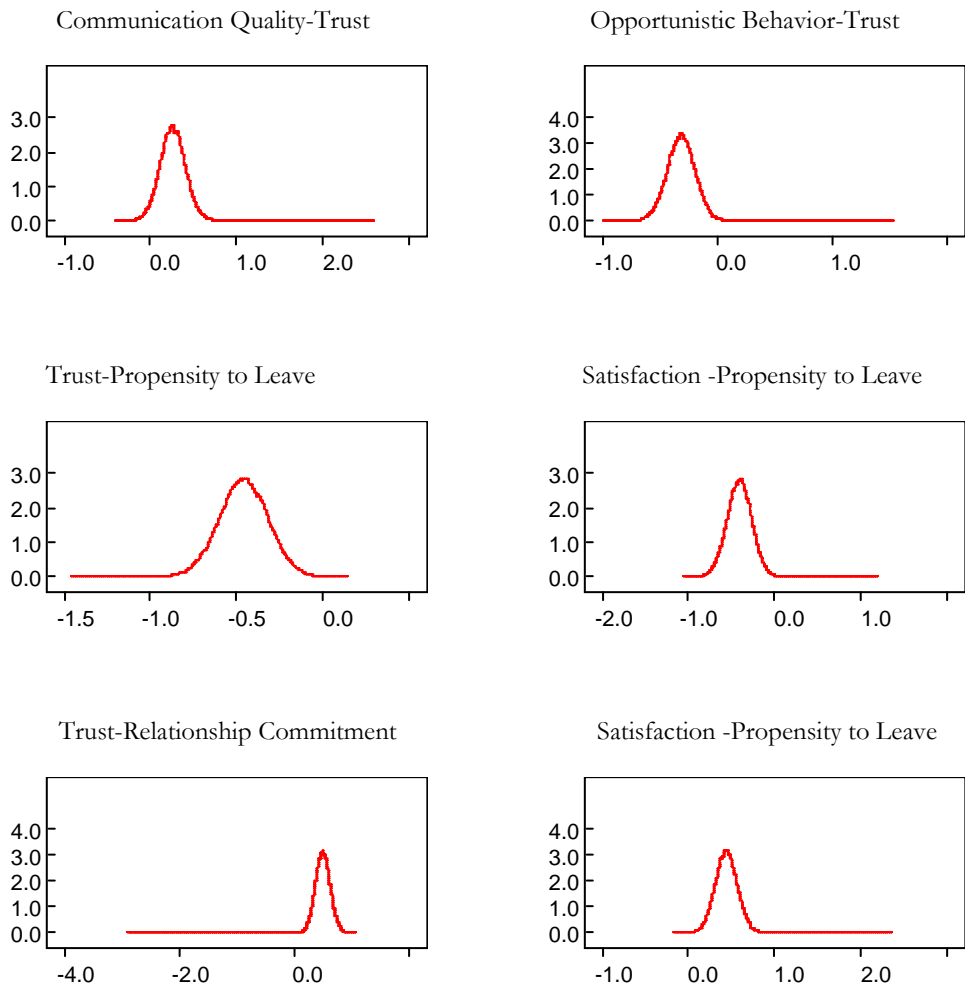


Table 4. Structural Model Results

Path	Mean	SD	95% <i>HPD Interval</i>	90% <i>HPD Interval</i>
Communication Quality → Trust	0.25	0.15	-0.04, 0.54	0.00, 0.49
Communication Quality → Relationship Satisfaction	0.22	0.15	-0.08, 0.52	-0.02, 0.46
Opportunistic Behavior → Trust	-0.30	0.12	-0.55, -0.06	-0.51, -0.10
Opportunistic Behavior → Relationship Satisfaction	-0.22	0.12	-0.47, 0.03	-0.43, -0.01
Financial Dependence → Trust	0.33	0.19	-0.01, 0.77	0.04, 0.67
Financial Dependence → Relationship Satisfaction	0.30	0.19	-0.04, 0.71	0.00, 0.62
Social Dependence → Trust	0.37	0.23	-0.02, 0.87	0.02, 0.78
Social Dependence → Relationship Satisfaction	0.39	0.23	-0.01, 0.88	0.05, 0.80
Trust → Propensity to Leave	-0.43	0.13	-0.75, -0.11	-0.70, -0.16
Trust → Relationship Commitment	0.49	0.14	0.22, 0.78	0.26, 0.73
Relationship Satisfaction → Propensity to Leave	-0.49	0.18	-0.87, -0.13	-0.81, -0.19
Relationship Satisfaction → Relationship Commitment	0.49	0.15	0.20, 0.80	0.25, 0.75

HPD stands for higher posterior density

Figure 4. Posterior Densities of Some Model Parameters





## APPENDIX 1: Full Description of a Bayesian Linear SEM

In line with Bayes' theorem in (9), we can write for example the posterior distribution of the SEM in (2) as follows:

$$\begin{aligned}
 & p(\Lambda_y, \Lambda_x, B, \Gamma, \Theta_\varepsilon, \Theta_\delta, \Psi, \Phi | Y) \propto \\
 & |\Theta_\varepsilon|^{-N/2} e^{\frac{1}{2} \sum_{i=1}^N (y_i - \Lambda_y \eta_i)' \Theta_\varepsilon^{-1} (y_i - \Lambda_y \eta_i)} |\Theta_\delta|^{-N/2} e^{\frac{1}{2} \sum_{i=1}^N (x_i - \Lambda_x \xi_i)' \Theta_\delta^{-1} (x_i - \Lambda_x \xi_i)} \\
 & |\Psi|^{-N/2} ||I - B||^{N/2} e^{-\frac{1}{2} \sum_{i=1}^N [(I-B)\eta_i - \Gamma \xi_i]' \Psi^{-1} [(I-B)\eta_i - \Gamma \xi_i]} \\
 & |\Phi|^{-N/2} e^{-\frac{1}{2} \sum_{i=1}^N \xi_i' \Phi^{-1} \xi_i} p(\Lambda_y, \Lambda_x, B, \Gamma, \Theta_\varepsilon, \Theta_\delta, \Psi, \Phi),
 \end{aligned} \tag{A1.1}$$

where  $p(\Lambda_y, \Lambda_x, B, \Gamma, \Theta_\varepsilon, \Theta_\delta, \Psi, \Phi)$  denotes the prior and the rest denotes the likelihood. The term  $||I - B||^{N/2}$  is the so-called Jacobian of transformation that we need when we have to find the joint distribution of  $\eta_i$  from the distribution of  $\zeta_i$  given  $B, \Gamma$  and  $\xi_i$ ). To proceed, first we write (1) in the form

$$z_i = \Lambda \omega_i + \varepsilon_i^z, \varepsilon_i^z \sim N_{p+q} \left( 0, \Theta \right),$$

where  $z_i = \begin{bmatrix} y_i \\ x_i \end{bmatrix}, \Lambda = \begin{bmatrix} \Lambda_y & \\ & \Lambda_x \end{bmatrix}, \omega_i = \begin{bmatrix} \eta_i \\ \xi_i \end{bmatrix}, \varepsilon_i^z = \begin{bmatrix} \varepsilon_i \\ \delta_i \end{bmatrix}, \Theta = \begin{bmatrix} \Theta_\varepsilon & \\ & \Theta_\delta \end{bmatrix} = \Theta_{hh}$ .

We will indicate the type and nature of prior information, as need arises. Most common priors that can be used, here include normal and Wishart (multivariate versions of gamma) distributions because they are conditionally conjugate— that is they match the functional forms of the posterior conditional distributions

We can write each equation in the form:

$$z_{ih} = \lambda_h' \omega_{ih} + \varepsilon_{ih}^z, h = 1, \dots, p + q, \tag{A1.2}$$

where  $\lambda_h$  is the  $h^{\text{th}}$  row of matrix  $\Lambda$ . Apparently not all elements of  $\Lambda_y$  and  $\Lambda_x$  are nonzero and, therefore,  $\lambda_h$  contains only the nonzero elements with the corresponding elements in  $\omega_{ih}$ .

In this form it is clear that the conditional posterior distributions of  $\lambda_h$  is as follows:

$$\lambda_h | \cdot \sim N \left( \hat{\lambda}_h + \underline{\lambda}_h, \Theta_{hh} \left( \Omega_h' \Omega_h \right)^{-1} + \underline{V}_h \right), \tag{A1.3}$$

where  $\hat{\lambda}_h = \left( \Omega_h' \Omega_h \right)^{-1} \Omega_h' z_h$ ,  $\Omega_h = [\omega_{ih}]$  after deleting the columns which correspond to zero elements of  $\lambda_h$ , and  $\underline{\lambda}_h, \underline{V}_h$  denote the prior mean and covariance matrix of  $\lambda_h$ .

The posterior conditional of  $\Theta_{hh}$  is:

$$\frac{Q + \sum_{i=1}^N (z_{ih} - \lambda'_h \omega_{ih})^2}{\Theta_{hh}} \mid \cdot \sim \chi^2_{N+Q}, \quad (\text{A.4})$$

where  $\underline{N}, \underline{Q}$  are prior parameters and  $\chi^2_{N+Q}$  denotes the chi-square distribution with the indicated degrees of freedom.

To derive the posterior conditional distributions of  $B, \Gamma$  we first consider their joint posterior conditional distribution:

$$p(B, \Gamma \mid \cdot, Y, X) \propto |\Psi|^{-N/2} \|I - B\|^{N/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (\eta_i - B\eta_i - \Gamma\xi_i)' \Psi^{-1} (\eta_i - B\eta_i - \Gamma\xi_i) \right\} p(B, \Gamma), \quad (\text{A.5})$$

where  $p(B, \Gamma)$  is a prior and the term  $\|I - B\|^{N/2}$  is the so-called Jacobian of transformation that we need when we need to find the joint distribution of  $\eta_i$  from the distribution of  $\zeta_i$  (given  $B, \Gamma$  and  $\xi_i$ ). Suppose  $\beta$  and  $\gamma$  are vectors that denote the non-zero elements of  $B$  and  $\Gamma$  respectively. Suppose the priors are

$$\beta \sim N(\underline{\beta}, V_\beta), \gamma \sim N(\underline{\gamma}, V_\gamma).$$

Conditionally on the nonzero elements of  $B$ , we can write the posterior conditional distribution of  $\Gamma$  as follows:

$$p(\Gamma \mid \cdot, Y, X) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (\psi_i - \Gamma\xi_i)' \Psi^{-1} (\psi_i - \Gamma\xi_i) \right\} p(\gamma), \quad (\text{A.6})$$

where  $\psi_i = (I - B)\eta_i$ . This is in the form of a multivariate regression model whose analysis has been taken up in detail by Zellner (1971, pp. 224-233).

The posterior conditional distribution of  $B$  is:

$$p(B \mid \cdot, Y, X) \propto |\Psi|^{-N/2} \|I - B\|^{N/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (\eta_i^* - B\eta_i)' \Psi^{-1} (\eta_i^* - B\eta_i) \right\} p(\beta), \quad (\text{A.7})$$

where  $\eta_i^* = \eta_i - \Gamma\xi_i$ . Due to the presence of the Jacobian term this distribution is not in any known family and, therefore, we use a Metropolis-Hastings algorithm to provide random draws.

The posterior conditional distribution of  $\Psi$  is:

$$p(\Psi | \cdot, Y, X) \propto |\Psi|^{-(N+M)/2} \exp\left\{-\frac{1}{2}\Psi^{-1}\text{tr}(S + \underline{S})\right\}, \quad (\text{A.8})$$

where  $S = \sum_{i=1}^N (\eta_i^* - B\eta_i)(\eta_i^* - B\eta_i)'$  and  $\underline{N}, \underline{S} = \underline{h}I$  are prior parameters. In particular,  $\underline{S}$  is a diagonal matrix and we set its parameter to  $\underline{h} = 0.001$  along with  $\underline{N} = 1$ , choices that impose minimal prior information on  $\Psi$ . This is in the form of a Wishart distribution, see Zellner (1971, pp. 389-390).

We next turn attention to the posterior conditional distributions of the latent variables. Regarding  $\xi_i$  we have the following sources of information:

$$\begin{aligned} x_i &= \Lambda_x \xi_i + \delta_i, \\ \psi_i &= \Gamma \xi_i + \zeta_i, \\ \xi_i &\sim N(0, \Phi). \end{aligned}$$

Therefore, the posterior conditional distribution is:

$$\xi_i | \cdot, Y, X \sim N(\hat{\xi}_i, V_\xi), \quad (33)$$

where

$$\begin{aligned} \hat{\xi}_i &= \left(\Lambda_x' \Theta_\delta^{-1} \Lambda_x + \Gamma' \Psi^{-1} \Gamma + \Phi^{-1}\right)^{-1} \left(\Lambda_x' \Theta_\delta^{-1} x_i + \Gamma' \Psi^{-1} \psi_i\right), \\ V_\xi &= \left(\Lambda_x' \Theta_\delta^{-1} \Lambda_x + \Gamma' \Psi^{-1} \Gamma + \Phi^{-1}\right)^{-1}. \end{aligned}$$

To understand this result we must remind the reader the following result known as Theil's mixed estimation. Suppose we have a linear model of the form

$$y = X\beta + u, \quad u \sim N(0, \Omega),$$

and prior information has the form

$$\beta \sim N(\bar{\beta}, V),$$

which by a “fiducial” argument<sup>13</sup> we can write as:

---

<sup>13</sup> Fiducial inference originally due to Fisher is an approach that starts from the Bayesian results and solves in terms of certain quantities of interest, for example sample mean, sample variance, regression coefficients. We do not wish to go into the intricate details here.

$$\bar{\beta} = I\beta + e, \quad e \sim N(0, V).$$

Here, we have two pieces of information which we can write jointly as:

$$\begin{bmatrix} y \\ \bar{\beta} \end{bmatrix} = \begin{bmatrix} X \\ I \end{bmatrix} \beta + \begin{bmatrix} u \\ e \end{bmatrix}.$$

The Generalized Least Squares (GLS) estimator in this model is:

$$\hat{\beta}_{GLS} = \left( \begin{bmatrix} X' & \mathbf{I} \end{bmatrix} \begin{bmatrix} \Omega^{-1} & \\ & V^{-1} \end{bmatrix} \begin{bmatrix} X \\ I \end{bmatrix} \right)^{-1} \begin{bmatrix} X' & \mathbf{I} \end{bmatrix} \begin{bmatrix} \Omega^{-1} & \\ & V^{-1} \end{bmatrix} \begin{bmatrix} y \\ \bar{\beta} \end{bmatrix} = \left( X' \Omega^{-1} X + V^{-1} \right)^{-1} \left( X' \Omega^{-1} y + V^{-1} \bar{\beta} \right). \quad (\text{A.9})$$

Theil called this the ‘‘mixed estimator’’ (when  $\Omega = \sigma^2 I$ ) but it turns out to be the mean of  $\beta$  (conditionally on  $\Omega$ ) in the Bayesian analysis of the linear model. The conditional covariance matrix of  $\beta$  is simply  $\text{cov}(\hat{\beta}_{GLS}) = (X' \Omega^{-1} X + V^{-1})^{-1}$ .

Regarding  $\eta_i$  we have the following sources of information:

$$\begin{aligned} y_i &= \Lambda_y \eta_i + \varepsilon_i, \\ (I - B) \eta_i &= \Gamma \xi_i + \zeta_i, \end{aligned}$$

from which we have:

(A.10)

$$\eta_i \mid \cdot, Y, X \sim N(\hat{\eta}_i, V_\eta),$$

where

$$\begin{aligned} \hat{\eta}_i &= \left[ \Lambda_y' \Theta_\varepsilon^{-1} \Lambda_y + (I - B)' \Psi^{-1} (I - B) \right]^{-1} \left[ \Lambda_y' \Theta_\varepsilon^{-1} y_i + (I - B)' \Psi^{-1} \Gamma \xi_i \right], \\ V_\eta &= \left[ \Lambda_y' \Theta_\varepsilon^{-1} \Lambda_y + (I - B)' \Psi^{-1} (I - B) \right]^{-1}. \end{aligned}$$

## APPENDIX 2: Winbugs Code for Application 1

```

model {
  for(i in 1:N){
    #measurement equation model
    for(j in 1:P){
      y[i,j]~dnorm(mu[i,j],psi[j])
      e[i,j]<-y[i,j]-mu[i,j]
      r[i,j]<-e[i,j]/y[i,j]
    }

    #Communication Quality
    mu[i,1]<-xi[i,1]+alp[1]
    mu[i,2]<-lam[1]*xi[i,1]+alp[2]
    mu[i,3]<-lam[2]*xi[i,1]+alp[3]
    mu[i,4]<-lam[3]*xi[i,1]+alp[4]
    #Opportunistic Behavior
    mu[i,5]<-xi[i,2]+alp[5]
    mu[i,6]<-lam[4]*xi[i,2]+alp[6]
    mu[i,7]<-lam[5]*xi[i,2]+alp[7]
    mu[i,8]<-lam[6]*xi[i,2]+alp[8]
    #Financial Dependence
    mu[i,9]<-xi[i,3]+alp[9]
    mu[i,10]<-lam[7]*xi[i,3]+alp[10]
    #Social Dependence
    mu[i,11]<-xi[i,4]+alp[11]
    mu[i,12]<-lam[8]*xi[i,4]+alp[12]
    #Trust
    mu[i,13]<-xi[i,5]+alp[13]
    mu[i,14]<-lam[9]*xi[i,5]+alp[14]
    mu[i,15]<-lam[10]*xi[i,5]+alp[15]
    mu[i,16]<-lam[11]*xi[i,5]+alp[16]
    mu[i,17]<-lam[12]*xi[i,5]+alp[17]
    # Relationship Satisfaction
    mu[i,18]<-xi[i,6]+alp[18]
    mu[i,19]<-lam[13]*xi[i,6]+alp[19]
    mu[i,20]<-lam[14]*xi[i,6]+alp[20]
    # Propensity to Leave
    mu[i,21]<-eta[i,1]+alp[21]
    mu[i,22]<-lam[15]*eta[i,1]+alp[22]
    mu[i,23]<-lam[16]*eta[i,1]+alp[23]
    # Relationship Commitment
    mu[i,24]<-eta[i,2]+alp[24]
    mu[i,25]<-lam[17]*eta[i,2]+alp[25]
    mu[i,26]<-lam[18]*eta[i,2]+alp[26]
    mu[i,27]<-lam[19]*eta[i,2]+alp[27]

    #structural equation model
    xi[i,1:4]~dmnorm(u[1:4],phi[1:4,1:4])
    xi[i,5]~dnorm(nu1[i],psd)
    nu1[i]<-gama1*xi[i,1]+gama2*xi[i,2]+gama3*xi[i,3]+gama4*xi[i,4]
    xi[i,6]~dnorm(nu2[i],psd)
    nu2[i]<-alpha1*xi[i,1]+alpha2*xi[i,2]+alpha3*xi[i,3]+alpha4*xi[i,4]
    eta[i,1]~dnorm(nu3[i],psd)
    nu3[i]<-eta1*xi[i,5]+eta2*xi[i,6]
    eta[i,2]~dnorm(nu4[i],psd)
    nu4[i]<-theta1*xi[i,5]+theta2*xi[i,6]
  }
}

```

```

} #end of i

#priors on intercepts
  #priors on intercepts
for(j in 1:27) {alp[j]~dnorm(0.03, 1)}

#priors on loadings and coefficients
lam[1]~dnorm(0.9,psi[2])
.
gama1~dnorm(0.9, psd)
gama2~dnorm(0.9, psd)
gama3~dnorm(0.9, psd)
gama4~dnorm(0.9, psd)
alpha1~dnorm(0.9, psd)
alpha2~dnorm(0.9, psd)
alpha3~dnorm(0.9, psd)
alpha4~dnorm(0.9, psd)
eta1~dnorm(0.9, psd)
eta2~dnorm(0.9, psd)
theta1~dnorm(0.9, psd)
theta2~dnorm(0.9, psd)

#priors on precisions
for(j in 1:P) {
  psi[j]~dgamma(10, 10)
  sgm[j]<-1/psi[j]
}
psd~dgamma(10, 1)
sgd<-1/psd
phi[1:4,1:4]~dwish(R[1:4,1:4],4)
phx[1:4,1:4]<-inverse(phi[1:4,1:4])

} #end of model

```

### APPENDIX 3: Markov Chain Monte Carlo

We use a Girolami and Calderhead (2011, GC) algorithm to update draws for a parameter  $\theta$ . The algorithm uses local information about both the gradient and the Hessian of the log-posterior conditional of  $\theta$  at the existing draw. A Metropolis test is again used for accepting the candidate so generated but the GC algorithm moves considerably faster relative to our naive scheme previously described. The GC algorithm is started at the first-stage GMM estimator and MCMC is run until convergence. It has been found that the GC algorithm performs vastly superior relative to the standard Metropolis-Hastings algorithm and autocorrelations are much smaller.

Suppose  $L(\theta) = \log p(\theta | \mathcal{D})$  is used to denote for simplicity the log posterior of  $\theta$ . Moreover, define:

$$\mathbf{G}(\theta) = \text{est.cov} \frac{\partial}{\partial \theta} \log p(\mathcal{D} | \theta), \quad (\text{A3.1})$$

the empirical counterpart of:

$$\mathbf{G}_o(\theta) = -E_{Y|\theta} \frac{\partial^2}{\partial \theta \partial \theta'} \log p(\mathcal{D} | \theta) \quad (\text{A3.2})$$

The Langevin diffusion is given by the following stochastic differential equation:

$$d\theta(t) = \frac{1}{2} \tilde{\nabla}_\theta L\{\theta(t)\} dt + d\mathbf{B}(t), \quad (\text{A3.3})$$

where

$$\tilde{\nabla}_\theta L\{\theta(t)\} = -\mathbf{G}^{-1}\{\theta(t)\} \cdot \nabla_\theta L\{\theta(t)\}, \quad (\text{A3.4})$$

is the so called “natural gradient” of the Riemann manifold generated by the log posterior. The elements of the Brownian motion are:

$$\begin{aligned} \mathbf{G}^{-1}\{\theta(t)\} d\mathbf{B}_i(t) = & |\mathbf{G}\{\theta(t)\}|^{-1/2} \sum_{j=1}^{K_\theta} \frac{\partial}{\partial \theta} \left[ \mathbf{G}^{-1}\{\theta(t)\}_{ij} |\mathbf{G}\{\theta(t)\}|^{1/2} \right] dt \\ & + \left[ \sqrt{\mathbf{G}\{\theta(t)\}} d\mathbf{B}(t) \right]_i. \end{aligned} \quad (\text{A3.5})$$

The discrete form of the stochastic differential equation provides a proposal as follows:

$$\begin{aligned} \tilde{\theta}_i = & \theta_i^o + \frac{\varepsilon^2}{2} \left\{ \mathbf{G}^{-1}(\theta^o) \nabla_\theta L(\theta^o) \right\}_i - \varepsilon^2 \sum_{j=1}^{K_\theta} \left\{ \mathbf{G}^{-1}(\theta^o) \frac{\partial \mathbf{G}(\theta^o)}{\partial \theta_j} \mathbf{G}^{-1}(\theta^o) \right\}_{ij} \\ & + \frac{\varepsilon^2}{2} \sum_{j=1}^{K_\theta} \left\{ \mathbf{G}^{-1}(\theta^o) \right\}_{ij} \text{tr} \left\{ \mathbf{G}^{-1}(\theta^o) \frac{\partial \mathbf{G}(\mathbf{a}^o)}{\partial \theta_j} \right\} + \left\{ \varepsilon \sqrt{\mathbf{G}^{-1}(\theta^o)} \xi^o \right\}_i \\ = & \mu(\theta^o, \varepsilon)_i + \left\{ \varepsilon \sqrt{\mathbf{G}^{-1}(\theta^o)} \xi^o \right\}_i, \end{aligned}$$

where  $\beta^o$  is the current draw and  $\varepsilon$  is a constant determined by trial-and-error so that approximately  $1/4$  of all draws are eventually accepted. The proposal density is:

$$q(\tilde{\theta} | \theta^o) = N_{k_\theta}(\tilde{\theta}, \varepsilon^2 \mathbf{G}^{-1}(\theta^o)), \quad (\text{A3.6})$$

and convergence to the invariant distribution is ensured by using the standard form Metropolis-Hastings probability:

$$\min \left\{ 1, \frac{p(\tilde{\theta} | \cdot, \mathcal{D}) q(\theta^o | \tilde{\theta})}{p(\theta^o | \cdot, \mathcal{D}) q(\tilde{\theta} | \mathbf{a}^o)} \right\}. \quad (\text{A3.7})$$