

Unobserved Heterogeneity in Hospitality and Tourism Research

Albert G. Assaf
Associate Professor
Isenberg school of Management
University of Massachusetts-Amherst
Email: assaf@isenberg.umass.edu

Haemoon Oh
Professor
Isenberg school of Management
University of Massachusetts-Amherst
Email: Oh@isenberg.umass.edu

Mike Tsionas
Professor
Lancaster University Management School
Email: m.tsionas@lancaster.ac.uk

Abstract

Despite the growing complexity of SEM applications in tourism, it is surprising that most applications have estimated these models without accounting for unobserved heterogeneity. In this paper we aim to discuss the concept of unobserved heterogeneity in more detail, highlighting its serious threats to the validity and reliability of SEMs. We describe a Bayesian finite mixture modeling framework for estimating SEMs while accounting for unobserved heterogeneity. We provide a comprehensive description of this model, and provide guidance on its estimation using the Winbugs software. We illustrate the importance of unobserved heterogeneity and the finite mixture modeling framework using a didactic application on brand equity where heterogeneity is likely to play an important role due to the differences in how consumers perceive the different dimensions of brand equity. We compare between various models and illustrate the differences between the standard and heterogeneous SEM, and discuss the implications for research and practice.

1. INTRODUCTION

Over the last decade, structural equation models (SEMs) have been increasingly popular across many tourism disciplines (Nunkoo et al. 2013; do Valle and Assaker, 2015), such as sustainable tourism (Zhang and Lei, 2012), tourism marketing (Chi and Qu, 2008), and tourism management (Yoon and Uysal, 2005; Hallmann et al. 2015). These models are highly flexible and powerful in the sense that they allow researchers to model and test complex patterns of relationships between latent and observed variables that would not be possible using other methods (Reisinger and Turner, 1999). Though progress continues to be made in the application of these models in tourism, most studies are still often simplistic and based on the implicit assumption that the sample is drawn from a single population.

Several researchers have warned about the issue of estimating homogenous models when testing, for example, theories of customer satisfaction, as customer decisions may vary across different segments or groups (Fonseca, 2009; Sarstedt and Ringle, 2010). As emphasized by Becker et al. (2013), ignoring unobserved heterogeneity across various groups can bias parameter estimates, resulting in incorrect hypothesis testing. For example, while a homogenous SEM may reveal a strong goodness-of-fit index, leaving the researcher satisfied that the model is valid, the parameters of the model might still be severely biased if unobserved heterogeneity exists and not properly recognized in the estimation. The seminal paper by Jedidi et al. (1997), for instance, emphasized this issue with an application from marketing, where they compared between two models with one that accounted for unobserved heterogeneity and the other that failed to recognize unobserved heterogeneity. They showed that while the fit statistics were perfect for the homogeneous model, the results were meaningless and the structural parameters were severely biased. More seriously, they emphasized that the “traditional fit statistics do not provide diagnostic information alerting the researcher to the presence of unaccounted unobserved heterogeneity in the sample” (Jedidi et al. 1997, p. 40). In addition, there is a serious threat to major types of model validity such as internal and instrumental validity, as illustrated by recent simulation evidence (Becker et al. 2013). Moreover, as ignoring heterogeneity can result in biases and inconsistencies in all

parameters estimates, the effect of covariates or mediating variables cannot also be meaningful.

While all these issues are now well established in marketing and management, it is surprising that the topic is yet to gain strong foothold in tourism research. Though recent tourism models have acknowledged the importance of unobserved heterogeneity (Mazanec and Strasser, 2007), the focus in tourism tends to be more on observed rather than unobserved heterogeneity. The two concepts are different and should not be confused each other. Observed heterogeneity refers to a situation when the researcher has a priori theoretical knowledge about the group differences in the data (Becker et al. 2013). This happens, for example, when a tourism researcher includes particular moderators (e.g. age, income level, etc.) in research design to account for heterogeneity for the various relationships in SEM. Unobserved heterogeneity, in contrast, aims to capture situations when there is no clear theoretical account for heterogeneity in the data, despite potentially being present in the data (Becker et al. 2013). More specifically, unobserved heterogeneity mainly refers to either latent class membership¹ or a violation of the usual iid assumption, i.e. that an observed variable is drawn not from a single population probability density function (PDF), but multiple PDFs and not observable variation, easily measurable by variance or IQR, or by staring at a histogram or boxplot. As an example in a panel data model $y_{it} = \alpha_i + \beta x_{it} + u_{it}$, the variation due to x_{it} is observed heterogeneity while variation in α_i stands for unobserved heterogeneity. Unobserved heterogeneity can take the form of membership into different population PDF which is, for example, the case when an observation y_i can belong to a population $N(x_i\beta_1, \sigma_1^2)$ with probability π and a population $N(x_i\beta_2, \sigma_2^2)$ with probability $1-\pi$.

The aim of this paper is to describe a modelling framework for accounting for unobserved heterogeneity in SEM tourism applications. We focus on (1) describing the heterogeneity concept in more detail, (2) presenting a finite mixture structural equation model (FMSEM) to handle unobserved heterogeneity in SEMs, (3) discussing how to test and identify the correct number of groups in the FMSEM by using the Bayesian approach. As the tourism literature

¹ Note that latent class analysis is just another method of cluster analysis, with the same core conceptual basis: clusters are not “revealed”; they are constructed.

currently lacks a clear description of this model, we see possible merits in this approach for structural equation modeling in tourism research. The focus on how to estimate finite mixture models in a Bayesian framework is particularly important as the Bayesian approach introduces more flexibility in the estimation of more advanced version of finite mixture models.

We discuss in this paper two versions of the FMSEM model including a standard FMSEM and an extended FMSEM. The latter allows the inclusion of covariates in the mixing probability, providing therefore more meaningful interpretation of the sources of heterogeneity between groups. We illustrate the model with an application from the tourism literature, and provide researchers with clear instructions on how to estimate the model using the Bayesian Winbugs software. In the following, section 2 discusses in more detail the concept of heterogeneity in SEM. Section 3 introduces the Bayesian finite mixture model. Section 4 presents an application example. Section 5 tests the robustness of the model. Finally, sections 6, 7 and 8 discuss the results, implications of our study, and suggested avenues for future research.

2. OBSERVED vs. UNOBSERVED HETERGONEITY

As mentioned, in the discussion of heterogeneity it is important to distinguish between observed and unobserved heterogeneity. The first reflects a situation where differences between groups or segments are expected *a priori*, and can be justified by incorporating individual-specific variables or moderators in the model (Becker et al. 2013). Studies in tourism, for example, have incorporated moderators such as “consumer product involvement” to explain the satisfaction-loyalty relationship (Chen and Tsai, 2008), “quality” to explain the perceived value-satisfaction relationship (Petrick, 2004), and “motivations” to explain the choice of tourism destinations (Nicolau and Mas, 2006). Observed heterogeneity is also commonly explained using demographic (Matzler et al. 2008) income (Barros and Machado, 2010), education (Funk and Bruun, 2007) or time variables (Kim et al. 2013).

Unobserved heterogeneity reflects a situation when heterogeneity exists in the model though not explicitly defined, or when the moderators or individual-specific variables in the model do not fully account for the sources of heterogeneity in the population (Becker et al. 2013). In such situations, researchers need to use appropriate modelling techniques to account for

unobserved heterogeneity and identify the number of segments or groups in the data (Hutchinson, Kamakura and Lynch, 2000; Horsky, Misra and Nelson, 2006). Such process can then be followed by a post hoc analysis to provide theoretical support for the grouping. Of course, in situations when theoretical explanations are not available to transform unobserved heterogeneity into observed heterogeneity, a researcher may use the number of identified groups to provide new theoretical directions to a certain framework (Becker et al. 2013).

There are, for example, many situations when a researcher started with a certain theoretical belief about a model and later proven wrong as a result of accounting for unobserved heterogeneity (Ansari et al. 2000). The literature provides now enough evidence about the serious threats of ignoring unobserved heterogeneity not only in the context of SEMs but also in other modeling frameworks such as regressions or panel data models (Kahn and Raftery, 1996; Wooldridge, 2005). For SEMs the risk is two-sided as ignoring heterogeneity has an effect on both the measurement and structural parts of the model (Becker et al. 2013). Simulation evidence, for example, has shown that ignoring unobserved heterogeneity affects the sign and size of the structural parameters, provides misleading conclusions about the strength of the relationships, and decreases the overall power of the model (Jedidi et al. 1997; Lee and Song, 2012; Becker et al. 2013). For the measurement side, ignoring heterogeneity can also bias the results for both formative and reflective models. With reflective models, for example, ignoring heterogeneity can affect the loadings as respondents “may provide information with different degrees of accuracy (e.g., variability in measurement error for any given scale item could differ across individuals)” (Ansari et al. 2000, p.329). In addition, when the construct’s measures are formative, ignoring heterogeneity can affect the indicators’ weights across groups leading to wrong interpretation about the importance of each of the different facets loadings into the main construct.

The paper by Becker et al. (2013) provided a useful example on this issue using the famous service quality (SERVQUAL) model that was measured through these formative indicators: (1) “tangibles”, (2) “reliability”, (3) “assurance”, (4) “empathy”, and (5) “responsiveness”

(Parasuraman et al. 1985). In assessing this model, customers might be heterogeneous in the answers, as some might prefer the communication facets such as empathy and responsiveness, while others might prefer the trust facets such as assurance and reliability. Such heterogeneity in customers' preferences might lead to different weights across the groups and, if not handled correctly, would lead to misleading conclusions. When pooling the sample, for example, a researcher is simply treating the two groups as equally important in their evaluation of SERVQUAL, ignoring the fact that this overall model does not represent either groups. In addition, because of this, the relationship between SERVQUAL and other potential constructs in the model such as customer satisfaction is also likely to be biased, not to mention that ignoring heterogeneity can also affect different types of SEM model validity such as internal validity, statistical validity and construct validity. The threat to internal validity, for example, happens because not accounting for unobserved heterogeneity can hide some new segments in the data that are not fully covered by the moderators or other individual specific variables, while the threat to statistical validity happens because ignoring unobserved heterogeneity can affect the size and sign of the coefficients and increase the standard error of the estimates. Finally, the threat to construct validity can result from not detecting the true group-specific measures of the constructs (see Becker et al. 2013 for more detailed discussion of these issues).

While studies in marketing and management have started to pay strong attention to the above issues, the measurement of unobserved heterogeneity is still largely ignored in the tourism literature. Most theories tested in tourism (e.g. SERVQUAL, customer satisfaction) are complex and involve customer surveys and opinion on different constructs, which can be subject to unobserved heterogeneity. Often we see, for example, meta-analysis in tourism referring to conflicting findings due to the absence of some moderators or contextual variables from some studies (e.g. Weed, 2006; Stepchenkova and Mills, 2010). However, accounting for such moderators and other contextual variables can only address the observed heterogeneity part of the model. Tourism researchers need to be aware that conflicting results between studies that focus on the same theory can also result from unobserved heterogeneity. In other words,

unobserved heterogeneity can introduce new dimensions to existing theories that would help unveil some of the inconsistencies in the literature.

3. APPROACHES TO ACCOUNT FOR HETEROGENEITY

The literature offers several methods to handle observed and unobserved heterogeneity. As mentioned, if heterogeneity can be easily observed or defined *a priori* it is common to use moderators or other contextual factors to differentiate between various groups in the data. This approach for instance, is common in the tourism literature (Ekinci and Hosani, 2006; Ferrer-Rosell et al. 2014). Alternatively one can split the sample into groups and the data analysed with methods for multiple groups. Tourism researchers, for example, have commonly used multiple-group analyses or experimental controlled groups to compare explicitly defined groups such as gender groups, age groups, etc. (Jurovski and Gursoy, 2004; Okazaki and Hirose, 2009; Oh and Hsu, 2014). Such approaches, however, are not appropriate in the context of unobserved heterogeneity. Here “the variables that cause the heterogeneity in the data are not known beforehand. Consequently, it is also not known to which of the subpopulations a participant belongs, and it is not possible to divide the sample into groups” (Lubke and Muthen, 2005, p. 21). As the potential groups are not directly observed (i.e. latent) and have to be derived or inferred from the data (Jedidi et al. 1997; Lubke and Muthen, 2005), one simple solution to the problem would be to use a two-stage procedure combining cluster analysis in the first with multi-group analysis in the second. As theoretical and simulation evidence (Jedidi et al. 1997; Görz, Hildebrandt, & Anacker, 2000), however, have proved that such approach may provide biased conclusions, the recent literature has focused on two main methods to handle unobserved heterogeneity: the finite mixture SEM (FMSEM) and the Bayesian hierarchical SEM (BHSEM).

The FMSEM was initially introduced by Jedidi et al. (1997) in the context of market segmentation. These models generalize the multigroup SEM to the context where groups cannot be observed *a priori* (i.e. unobserved heterogeneity). The FMSEM is highly popular and has been tested successfully across several fields including behavioural, social, medical and environmental sciences. The model “arises with a population which is a mixture of K

components (latent classes) with probability densities ($\{f_k, k=1, \dots, K\}$) and mixing proportions $\{\pi_k, k=1, \dots, K\}$ ” (Lee and Song, 2012, p.162). Hence, FMSEM represents of heterogeneity in a finite number of components or latent classes. The difference between the FMSEM and the two-stage approach highlighted above is that the “FMSEM assigns the observations to a pre-specified number groups by means of fuzzy (probabilistic) clustering, thereby permitting the simultaneous estimation of group-specific parameters” (Becker et al., 2013, p. 673). Hence, this avoids the bias with the two stage approach.

The performance of the FMSEM has been tested in several studies (Temme et al. 2002). It is highly flexible in fitting models with skewness and non-standard distributional characteristics (Lee and Song, 2012). Simulation evidence and findings from related studies have also shown that using the FMSEM can prevent researchers from making wrong conclusions about the structural and measurement relationships in SEMs (Jedidi et al. 1997; Lee and Song, 2012).

The Bayesian hierarchical SEM (BHSEM) has also been extensively used to handle unobserved heterogeneity. This model, originally introduced by Ansari et al. (2000), provides an important extension on multilevel models (also known as random coefficient models). For instance, in contrast to multilevel models which can only handle heterogeneity in mean structure (i.e. factor means, measurement mean intercepts), this model can also handle heterogeneity in structural parameters and covariance structures (Jedidi and Ansari, 2001; Preacher et al. 2010). As in most marketing and management applications heterogeneity exists beyond mean structure, the BHSEM provides hence a more comprehensive and robust assessment of heterogeneity (Ansari et al. 2000). Its main limitations, however, is that can work only with continuous data (Becker et al. 2013). This model also requires panel data per individual as it measures heterogeneity at the individual level and not among a defined number of groups (Becker et al. 2013). In other words, it is more suitable for recursive SEMs.

Because of these limitations, the FMSEM has been used more extensively in the marketing and management literature. This model is more flexible and can be used for different types of data and applications (Lee and Song, 2003). Its statistical estimation is however not straightforward. Researchers for example have traditionally struggled to correctly identify the number of heterogeneous groups with the FMSEM (Nylund et al. 2007). While some solutions have become available, they mostly require more complicated estimation techniques such as the Bayesian approach. In addition, some modified versions of the FMSEM also necessitate more advanced estimation methods such as the Bayesian approach.

Our focus in this paper is to simplify the description of the FMSEM and provide tourism researchers with complete guidance on how to estimate this model. Though a variety of estimation methods have been proposed to estimate FMSEM (e.g. method of moments, maximum likelihood (ML), minimum chi-square and Bayesian approach), the present paper focuses on the Bayesian approach. There are indeed computational reasons to use the Bayesian approach² when it comes to mixture models. Classical techniques, for example ML, have a serious problem as the likelihood becomes unbounded when one variance goes to zero (variance = σ^2 of a certain component in the mixture). We fully understand that the Bayesian approach is yet to gain a strong foothold in tourism research. In addition, there has been a lack of studies describing the Bayesian approach in the tourism literature. For this reason, we provide here a short description of the approach in the context of FMSEM and provide tourism researchers with clear guidelines on how to estimate the FMSEM in a Bayesian approach.

² In particular, the Markov chain Monte Carlo (MCMC) in Bayesian inference.

4. THE MODEL

4.1. FINITE MIXTURE STRUCTURAL EQUATION MODEL (FMSEM)

To illustrate the difference between the FMSEM and the standard SEM³, we start first with a description of the latter. The measurement and structural parts of the standard SEM can be described in (1) and (2), respectively, as follows:

$$y = \mu + \Lambda\omega + \varepsilon \quad (1)$$

$$\eta = \Pi\eta + \Gamma\xi + \delta \quad (2)$$

where for (1), y is a random vector of observed indicators for ω , a vector of latent variables, μ is a vector of intercepts, Λ is an unknown matrix of loadings representing the relationship between y and ω , and ε is a vector of residuals. For (2), η and ξ are random vectors representing the outcome and explanatory variables for w , Π and Γ are vector of regression coefficients that represent the effect on η_j , and δ is a vector of residuals.

While this model is simple to estimate, its main limitation is that it assumes that the sample is drawn from a population with homogenous characteristics (Becker et al. 2013). The FMSEM accounts for unobserved heterogeneity in the data using more flexible representations of the structural relationships between the latent variables and the implied distributions of the observed variables. In standard SEM, for example, it is typically assumed that there exists a single probability density function (e.g. standard normal distribution) from which the data are generated. In the FMSEM, however, this assumption is relaxed and instead it is assumed that the data are generated from a finite mixture multivariate normal density function⁴. Such

³ The standard SEM is the common SEM specification used in the tourism literature.

⁴ Mixtures involve the assumption that the model is correct for a particular group of observations, that the coefficients of this group are different compared to those of other groups, and that the allocation of observations to groups is unknown a priori (Lee and Song, 2003).

mixture distribution provides a natural representation of heterogeneity by dividing the whole population into sub-populations and for each subpopulation, the data can be modeled by a parametric distribution.

For example, for the observed indicators y_i we can write the mixture distribution as follows:

$$f(y_i) = \sum_{k=1}^K \pi_k f_k(y_i | \mu_k, \theta_k), \quad i = 1, \dots, n \quad (3)$$

where K is the number of subpopulations or groups with different characteristics specified by parameter vector θ_k , π_k is the unknown mixing proportion or the probability of an individual belonging to the k th subpopulation or group with the constraint that $\sum_{k=1}^K \pi_k = 1$, and $f_k(y_i | \mu_k, \theta_k)$ is the conditional density given that the observation is from the k th group with unknown mean vector .

Hence, for the k th group we can now write the measurement and structural parts described in (1) and (2) as:

$$y_i = \mu_k + \Lambda_k \omega_i + \varepsilon_i \quad (4)$$

$$\eta_i = \Pi_k \eta_i + \Gamma_k \xi_i + \delta_i \quad (5)$$

where y_i , μ_k , Λ_k , ω_i , ε_i , η_i , Π_k , Γ_k , ξ_i and δ_i are as defined under (1) and (2) but here in terms of the k th group. Any of these parameters can be invariant across groups. However, “it is essential to assign a different μ_k in the measurement equation of each component in order to effectively analyze data from the heterogeneous population that differ by their mean vectors” (Lee and Song, 2012, p.164).

4.2. AN EXTENDED FMSEM

In the FMSEM literature, researchers have recently highlighted the importance of allowing the mixing probability in the FMSEM to vary with a vector some exogenous covariates in order to explain the sources of unobserved heterogeneity between groups (Yuan and Bentler, 2010). For example, while the above FMSEM is enough to account for unobserved heterogeneity, it does not provide a rich interpretation on why different groups exist in the data.

To illustrate, we can define π_k , the mixing probability in (3) as :

$$\pi_k = p(z_i = k | x_i) = \frac{\exp(\tau_k^T x_i)}{\sum_{j=1}^K \exp\{\tau_j^T x_i\}} \quad (6)$$

where x_i is a vector of covariates, z_i is a latent allocation of y_i , and τ_k is a vector of coefficients. Hence, in this model, the components probability of y_i and the mixing probability π_k are related to a vector of covariates. These covariates can play an extremely helpful role in explaining the sources of differences between the various groups. The model hence adds increased flexibility to the standard FMSEM and allows for better and more parsimonious identification and description of heterogeneity.

However, as this modified version of the FMSEM is more complicated and highly non-linear, studies have recommended estimating the model using a Bayesian approach (e.g. Nylund et al. 2007). The traditional estimation methods such as maximum likelihood (ML), for example may not provide a good fit in such contexts (Cai et al. 2010).

4.3. BAYESIAN ESTIMATION

While different estimation methods (e.g. method of moments, maximum likelihood, Bayesian) have been proposed to estimate the FMSEM, we use in this paper the Bayesian approach for four main reasons. First, the Bayesian approach is well suited to handling the model in (6). The incorporation of covariates in the mixing probability, for example, complicates the FMSEM considerably and thus we need well-mixing procedures to perform the multivariate integrations

required by Bayesian analysis. In more complex models like (6), where, for example, latent variables enter the mixing probability, finding the covariance matrix becomes impossible or, at best, questionable as it uses various approximations.

Second, the use of the Bayesian approach offers both theoretical and practical benefits over frequentist methods (e.g. maximum likelihood) when dealing with heterogeneity in SEMs (Li and Wang, 2010; Cai et al. 2010). From “a practical viewpoint, Bayesian methods allow the flexible incorporation of prior information about model parameters. In addition, Bayesian methods allow the estimation of individual-specific estimates while accounting for parameter uncertainty in such estimates. Specifically, in our modelling context, the Bayesian methodology provides individual-specific estimates of the factor scores, structural coefficients, and other model parameters” (Ansari et al. 2000, p. 330). From a statistical view point, Bayesian methods do not rely on asymptotic theory⁵, which is the foundation for sampling theory treatments and estimation of the SEM.

Third, as mentioned, the Bayesian approach has also proven to be more reliable in identifying the correct number of groups in the FMSEM (Li and Wang, 2010). When estimating the model with maximum likelihood it is common to use the likelihood ratio test to compare between several FMSEMs with different groups and then select the best model based on the value of this ratio. Recent evidence, however, have shown that the likelihood ratio test performs poorly in the context of FMSEM. The Bayesian approach offers different modelling comparison criteria such as the Bayes factor or Deviance Information Criterion (DIC) that can be used to compare models and that have all proven to be highly reliable in the context of the FMSEM (Cai et al. 2010). These modelling criteria are also known to yield minimum risk discrimination among models or between hypotheses (Kass & Raftery, 1995).

⁵ It is important to emphasize here that with the Bayesian approach, we use normality of certain disturbances or normality of certain priors but this does not mean we do it because of asymptotic theory. We do it because normality is convenient and factor nicely with the likelihood (Coelli et al. 2005). If we change the priors, we will not obtain different results in large samples. The issue is that assuming normality for disturbances is not the same as asymptotic theory. Asymptotic refers to convergence of certain estimators to a certain distribution (most often, the normal). But posterior distributions need not be normal.

Finally, the Bayesian approach delivers results (for example, about posterior group classification) that are fully robust to parameter uncertainty. Likelihood ratio and other tests, on the contrary, condition on the ML estimates ignoring statistical variations in the parameters⁶.

5. APPLICATION

To illustrate the importance of accounting for unobserved heterogeneity, we apply the model in (4), (5) and (6) to a brand equity model recently proposed in the lodging industry by Hsu et al. (2011). This model (Figure 1) suggests that “brand loyalty” affects positively “brand choice intention” (i.e., the traveller’s willingness to choose the same brand in the presence of competitor). The study defines the following five positive determinants of brand loyalty: Perceived quality, brand awareness, brand image, management trust and brand reliability. For detailed conceptual discussions, model specification, and measurement properties, refer to Hsu et al. (2011)

While the study by Hsu et al. (2011)⁷ has specified the model in (1) using a homogenous SEM, there are many reasons to believe that unobserved heterogeneity exists in such types of models. For example, studies in marketing have discussed the importance of accounting for unobserved heterogeneity when modelling brand-choice data, as consumers are heterogeneous in the way they respond to the dimensions of brand choice intention (Horsky et al. 2006). There are also differences in “the underlying preferences consumers have for various brands” (Horsky et al. 2006, p. 322). Hence a failure to account for such heterogeneity might lead to inconsistent and biased results. In the following section, we present our results for both the homogeneous SEM (i.e. estimated without any account of unobserved heterogeneity) and

⁶ Specifically, likelihood tests condition on a particular estimate (for example the maximum likelihood estimate). By so doing, the variation in the parameter is ignored. In Bayesian analysis there is the well-known saying of Jeffreys that often a hypothesis is rejected (by a p-value) because it did not predict what we do not observe. Moreover, in p-values one allows for sampling variation that is variation in data that we did not observe.

⁷ We thank the authors for making the data of this study available for our illustration in this study.

the FMSEM. We illustrate the differences in findings between the two models and discuss the implications for practice.

6. DATA AND RESULTS

The sample for the model in Figure 1 includes 1346 observations collected from domestic and foreign visitors staying at 32 Chinese hotel properties. Tables 1 presents the items used to measure each of the constructs in Figure 1. These items were defined based on focus groups conducted by the authors as well as previous studies in the literature. We confirmed that the measurement model fit the data well as indicated by Hsu et al. (2011). The final items and their descriptive statistics are reported in Table 1.

6.1. HOMOGENEOUS SEM

Before presenting the results from the FMSEM, and for purpose of comparison, we present first the result from the homogenous (i.e. standard) SEM. In order to be consistent with the FMSEM we also estimate this model in a Bayesian framework⁸. We rely on the Winbugs software, which is highly reliable for Bayesian estimation and has been used in many recent management and marketing studies (Chen, Fay and Wang, 2011; Rust and Verhoef, 2005). We checked for model fit using the Posterior Predictive p-value (PPP), which is highly common with Bayesian estimation and has also been used in the context of SEM (Muthen and Asparouhov, 2012). Specifically, the PPP “uses the estimated posterior distribution and evaluates how that posterior distribution and the model fit the data” (Asparouhov and Bengt Muthen, 2010, p. 20), with values close to 0.5 indicating strong fit (Lee and Song, 2012)⁹.

In our context the PPP was around 0.52, confirming that the model fits the data well. The parameter estimates for this model are illustrated in Figure 2. As it is clear, there is a positive and statistically significant relationship between each of the six structural paths and brand loyalty, confirming previous theoretical expectations (Hsu et al. 2011). The relationship

⁸ Note that it is common in the tourism literature to estimate this model by using the Maximum Likelihood method available in, for example, LISREL or AMOS.

⁹ For more details on how to calculate the PPP refer to Muthen and Asparouhov (2011).

between brand loyalty and brand choice intention is particularly strong confirming that travellers “who are loyal to a hotel brand should show a strong willingness to choose the same brand among many competitor brands available” (Hsu et al. 2011, p.85).

6.2. THE FMSEM

6.2.1. DETERMINING THE NUMBER OF GROUPS

As part of the FMSEM estimation, and after confirming the model fits the data well (PPP=0.51), we first conducted some model comparison to determine how well the model performed against the homogenous SEM. It was also crucial to identify the number of groups required for the estimation of the FMSEM. A challenging and still unresolved aspect in the estimation of FMSEM, for instance, is to determine the number of unobserved heterogeneous groups in the data. There is contradiction in the literature regarding the best modelling criteria to use for determining the number of groups, though there has been some important progress in this area. Nylund et al. (2007), for example, showed through a simulation study that the Bayesian Information Criteria (BIC) was a reliable and consistent tool for correctly identifying the number of groups in finite mixture models. Lee and Song (2012) illustrated the reliability and importance of the Deviance Information Criterion (DIC) in determining the number of groups when estimating the FMSEM¹⁰. The authors also emphasized the importance of using the Bayesian approach and the DIC particularly when estimating complicated versions of the FMSEM such as the one illustrated in (6). The Maximum Likelihood (ML) approach-and its related tests (e.g. likelihood ratio test) – for instance, usually encounters serious difficulties due to the non-linearity of the FMSEM.

The DIC generalizes the Akaike Information Criterion (AIC), and is a “devised selection criterion which combines the Bayesian measures of complexity level and of how well the model fits the data” (Cai et al., 2009, p. 1865). Usually models with lower DIC are considered a better fit. The

¹⁰ As highlighted by Richardson and Green (1997), the Bayesian approach in general is very suitable to the problem when K is unknown.

DIC results are illustrated in Table 2. Along with the homogenous SEM, we estimated two FMSEMs, one assuming two groups and the other assuming three groups in the data. We can see first that the DIC for both FMSEMs is lower than the homogenous SEM, with the two-group model being the lowest in DIC, suggesting that (1) FMSEM is a better fit than the homogenous SEM in our branding example and (2) unobserved heterogeneity indeed exists in the data. Between the two FMSEMs we can see that the model with two groups outperforms that with three groups, hence a mixture model with two groups should be chosen. Before proceeding further and in order to confirm that the model is behaving well and revealing the true number of groups, we also conducted Monte Carlo Simulation, for which the details are provided in Appendix 2. The results clearly indicate that the Bayesian approach performs well across different sample size and the model is revealing with high confidence the true number of groups even in small sample size.

6.2.2. TESTING EQUIVALENCE

Now that the number of groups has been determined, we conducted a test for equivalence to ensure that the groups can be meaningfully compared. For instance, it is common in the FMSEM literature (Zyphur, Chaturvedi, & Arvey 2008) to ensure that the items are measuring the same constructs across the groups. Such equivalence in measurement can be tested by constraining the same factor loading to be equal across the groups. If equivalence does not exist, comparing between the structural parameters is meaningless as the groups come from different populations. In other words, we need to guarantee that “the same construct is assessed in different groups and that this construct has the same metrics characteristics among groups” (Elosua, 2011, p. 404).

This paper tests the between-group measurement equivalence by using the following χ^2 statistic. Suppose, first, that $\lambda = \text{vec}[\Lambda_y]$ and that the dimensionality of the vector is d (i.e. the total number of elements in matrix Λ_y). In the case of FMSEM the correct notation is

$\lambda_{(g)} = \text{vec}[\Lambda_y^{(g)}]$ for a given group. In the case of two groups (which turns out to be best in our application), define $\delta = \lambda_{(1)} - \lambda_{(2)}$. The statistic:

$$q = \delta'V^{-1}\delta$$

where $V = \text{Cov}(\delta)$, will follow a χ^2 distribution. The null hypothesis of the test is that there is equivalence between the two groups. From the results, we found that that the p-value of the test is equal to 0.83. Hence, there is clear support for equivalence in the factor loadings.

Moreover as the construct in both groups is measured on the same scale and using the same observed variables, we then have a solid base to compare between the two groups in our data set.

6.2.3. GROUP COMPARISON

The Bayesian estimates for both FMSEM groups are illustrated in Figures 3 and 4. As expected, most variables relate positively, as in the original model. There are, however, important differences between the two groups in terms of the size and significance of some coefficients. For example, we can see that both “perceived quality” and “brand awareness” do not have a significant impact on brand loyalty in Group 1 (see the dotted arrows), while brand image does not have a significant impact on brand loyalty in Group 2 (see the dotted arrow). We can also observe important differences in the size of the loadings between the two groups and also between each group and the homogeneous SEM in the figures¹¹.

6.2.4. COVARIATES OF UNOBSERVED HETEROGENEITY

Before discussing further the above results, we report in Table 3 the parameter estimates of the covariates that entered the mixing probability of the model in (6). As explained in (6), the

¹¹ We further confirmed this issue by monitoring the difference between the loadings of the various models in Winbugs. In most cases we confirmed (based on the posterior mean and the standard deviation) that there are significant differences between these loadings.

inclusion of these covariates was essential to provide richer interpretations of the sources of heterogeneity in the data, and to provide more clear justifications on the differences that existed between the two groups. Inclusion of covariates, in general, helps not only account for the background conditions for the functioning of a particular theory, but also develops a richer understanding of the theory and its boundary conditions.

For the selection of covariates in the FMSEM context, the researcher has the option to choose either the variables that exist completely outside the model (e.g. age, income, etc.) or the exogenous variables that were already included in the model. For illustration purposes in the present context, we tried to assess the impact of all exogenous variables that existed in the model (i.e. perceived quality, brand awareness, brand image, management trust and brand reliability). Such inclusion of the exogenous variables as covariates is well justified in the FMSEM literature (Cai et al. 2010; Lee and Song, 2012) in that their theoretical role in the model logically supports their immediate, additional roles as explanatory variables for classifying the sample into one of the two groups according to the mixing probability.

The results summarized in Table 3 represent the differences between groups 1 and 2 in terms of each of these variables, based on group 2 as the reference group for comparisons. Of the five covariates, two appeared to significantly determine group membership characteristics, in addition to their roles as antecedents to brand loyalty. For example, we can see that the respondents belonging to group 1 tend to have higher perceptions of overall quality (i.e. perceived quality) than those classified into group 2. However, group 1 respondents seem to have significantly lower perceptions of the overall image (i.e., brand image) than group 2. There is no difference between the two groups in terms of the other three variables.

These results indicate that in this study perceived quality and brand image were the main determinants of the underlying heterogeneity between the two groups. This, in fact, reflects how the hotel market is structured along the level of product and service quality, as implied in the industrywide star or diamond ratings. That is particularly true as perceived quality often becomes a main source forming the hotel's brand image and, combined, they impact customer attitudes in a distinct manner (Homer, 2008; Wu, Yeh, & Hsiao, 2011). Given their close

relationship, both conceptually and empirically, perceived quality and brand image could forge a powerful synergy to detect the underlying, effective heterogeneity and drive the mixing probability of classifying the sample into one of the two groups. In contrast, such classification power was absent in the concepts of brand awareness, management trust, and brand reliability even if their role in the conceptual model in Figure 1 was significant each.

Some important theoretical and practical implications are in order for discussions about these findings and FSMSEM in general. As to theoretical implications, the discriminatory grouping power of the exogenous variables (and covariates in this study) like these could not be uncovered by merely examining the homogenous SEM model. Models based on a homogeneity assumption may undermine reliability and validity of study results to the extent which the unveiled heterogeneity influences parameter estimates of the model, thereby leading to weak power in hypothesis testing as well as inconsistent results in future applications. In addition, recognizing the unobserved heterogeneity could alter our understanding of the relationships among the model constructs significantly. For example, the group-specific estimation of the model in Figures 3 and 4 produced some revealing differences in parameter estimates (see perceived quality, brand awareness, and brand image, in particular). In reality, empirical representation of a phenomenon, especially that of tourist behaviour, is likely to be an inverse function of imposing a homogeneity assumption on a population. Such a “generalist” approach tends to pose threat to rich theoretical development over time. We argue that hospitality and tourism research generally needs more conceptual and empirical investigations into underlying mechanisms such as unobserved heterogeneity we illustrated in this paper for at least two reasons. One is that our better understanding of tourist behaviour, for example, is likely to hinge upon often unobservable patterns of individual decisions making. Cultivating such underlying driving decision forces and structuring them into highly interpretable knowledge systems are at the core of theory building and scientific progress. Another reason relates to generalizability of research findings. Over-generalization of findings is pervasive in the hospitality and tourism literature and it comes partly from the ignorance of unobserved heterogeneity, among others, in many behavioural phenomena. As we mentioned earlier, a growing use of methods such as SEM is likely to serve our needs to examine the underlying

tourist behaviour process; yet, we find that the majority of SEM applications to date is centred on the data of one sample. Given the global and cross-cultural bearings of tourist behaviours, we need more concerted efforts to account for such vagaries as unobserved heterogeneity resulting from individual as well as cultural differences (Oh & Hsu, 2014).

Practical implications of understanding underlying heterogeneity relate largely to effectiveness of business strategies and policies. An explicit modelling of latent heterogeneity can assist in not only building rich theories around the topic in future studies but also improving precision in strategy design and policy implementation. For instance, identification of the variables that drive the group heterogeneity benefits market segmentation and targeting studies immediately, allowing destination marketers to fashion their market strategies to the customer's needs more precisely. As the degree of customization increase in strategy design, organizations will be able achieve their business goals more effectively by avoiding misaligned resource allocations. Such effective strategy alignment may certainly derive from proper positioning of the destination to the target market segments defined by powerful underlying heterogeneity.

7. Conclusions

This paper presented a thorough discussion of observed and unobserved heterogeneity illustrating the difference between the two. While observed heterogeneity is well established in the tourism literature, research so far has largely ignored unobserved heterogeneity. This paper discussed how unobserved heterogeneity can bias the results of SEM and affect the validity and reliability of the estimates. We introduced and tested a Bayesian FMSEM for handling and interpreting unobserved heterogeneity. Mixture models are quite useful in terms of "improving fit" and have been proven to be highly reliable in dealing with unobserved heterogeneity. The model we introduced is also unique in the sense that it allows the mixture probability to depend on covariates, providing a better alternative for modeling heterogeneity that admits interpretation. Such a model can be used routinely in practical applications as it allows for an interpretable and coherent improvement of fit.

As argued above, the use of the Bayesian approach to estimate FMSEM represents also another key contribution to the tourism literature. While different versions of finite mixture models are available in some SEM software such as Mplus, those are mostly not available in a Bayesian framework and do not allow the mixture probability to depend on covariates. As mentioned, the use of the Bayesian approach is essential to handling the complexity of our proposed model. It also provides flexibility to use the DIC when comparing different models. DIC have advantages over ML ratio tests as they are known to yield minimum risk discrimination among models or between hypotheses (Kass & Raftery, 1995).

The paper provided detailed instructions on how to estimate FMSEM in a Bayesian framework, as the estimation of the model is not possible in a standard SEM software such as LISREL, AMOS or Mplus. We relied on the Winbugs software, which started to gain increased popularity across the marketing and management disciplines. Tourism researchers can apply the code used in this paper to estimate FMSEM in other research contexts or at least to test for unobserved heterogeneity. The code can also be simply updated to estimate the simple SEM or other simpler versions of FMSEM such as the one without covariates in the mixing probability.

We illustrated the importance of FMSEM by using an example from the hospitality branding literature. We showed that in the context of our application FMSEM outperforms the standard SEM. We also illustrated that estimating the model using standard SEM when unobserved heterogeneity is present can bias the structural relationships of SEM. We argue that tourism researchers should always keep unobserved heterogeneity in mind, particularly if there are some theoretical indications that data are heterogeneous. If available, researchers should also test the existing model for unobserved heterogeneity on data sets grounded in substantive theories. In other words, it would be wise to test FMSEM on other data sets where authors have already identified the number of groups based on some existing theories or previous literature. In this case, FMSEM can serve as a post hoc modeling framework to test the validity of the group identity in the data. Given its flexible, yet robust modeling capabilities and given the fact that the tourism phenomenon is replete with preferential as well as destination diversities, FMSEM awaits much more exciting and creative applications in this discipline.

References

- Ansari, Asim, Kamel Jedidi, and Sharan Jagpal. (2000). "A Hierarchical Bayesian Methodology for Treating Heterogeneity in Structural Equation Models." *Marketing Science*, 19 (4): 328-47.
- Barros, Carlos Pestana, and Luis Pinto Machado. (2010). "The Length of Stay in Tourism." *Annals of Tourism Research*, 37 (3): 692-706.
- Becker, Jan-Michael, Arun Rai, Christian M. Ringle, and Franziska Völckner. (2013). "Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats." *MIS Quarterly*, 37 (3): 665-94.
- Cai, Jing-Heng, Xin-Yuan Song, and Yih-Ing Hser. (2010). "A Bayesian Analysis of Mixture Structural Equation Models with Non-Ignorable Missing Responses and Covariates." *Statistics in Medicine*, 29 (18): 1861-74.
- Chen, Ching-Fu, and Meng-Huan Tsai. (2008). "Perceived Value, Satisfaction, and Loyalty of TV Travel Product Shopping: Involvement as a Moderator." *Tourism Management*, 29 (6): 1166-71.
- Chen, Yubo, Fay, Scott, and Wang, Qi. (2011). "The Role of Marketing in Social Media: How Online Consumer Reviews Evolve." *Journal of Interactive Marketing*, 25 (2): 85-94.
- Chi, Christina Geng-Qing, and Hailin Qu. (2008). "Examining the Structural Relationships of Destination Image, Tourist Satisfaction and Destination Loyalty: An Integrated Approach." *Tourism Management*, 29 (4): 624-36.
- Coelli, T. J, Rao, D. S. P, O'Donnell, C. J, and Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.
- Congdon, Peter. (2007). *Bayesian Statistical Modelling*. West Sussex: John Wiley & Sons.
- do Valle, P. O., & Assaker, G. (2015). Using Partial Least Squares Structural Equation Modeling in Tourism Research A Review of Past Research and Recommendations for Future Applications. *Journal of Travel Research*, in press.

- Ekinci, Yuksel, and Sameer Hosany. (2006). "Destination Personality: An Application of Brand Personality to Tourism Destinations." *Journal of Travel Research*, 45 (2): 127-39.
- Elosua, Paula. (2011). "Subjective Values of Quality of Life Dimensions in Elderly People. A SEM Preference Model Approach." *Social Indicators Research*, 104 (3): 427-37.
- Ferrer-Rosell, Berta, Esther Martínez-García, and Germà Coenders. (2014). "Package and No-Frills Air Carriers as Moderators of Length of Stay." *Tourism Management*, 42: 114-22.
- Fonseca, Jaime R.S. (2009). "Customer Satisfaction Study via a Latent Segment Model." *Journal of Retailing and Consumer Services*, 16 (5): 352-59.
- Funk, Daniel C., and Tennille J. Bruun. (2007). "The Role of Socio-Psychological and Culture-Education Motives in Marketing International Sport Tourism: A Cross-Cultural Perspective." *Tourism Management*, 28 (3): 806-19.
- Görz, Nicole, Lutz, Hildebrandt; and Annacker, Dirk. (2000). "Analyzing Multigroup Data with Structural Equation Models, in Gaul, W.; Decker, R. (Hrsg.) *Classification and Information Processing at the Turn of the Millennium*, S. 312-319.
- Hallmann, K., Zehrer, A., & Müller, S. (2015). Perceived Destination Image: An Image Model for a Winter Sports Destination and Its Effect on Intention to Revisit. *Journal of Travel Research*, in press.
- Homer, Pamela Miles. (2008). "Perceived Quality and Image: When All Is Not "Rosy"." *Journal of Business Research*, 61 (7): 715-23.
- Horsky, Dan, Sanjog Misra, and Paul Nelson. (2006). "Observed and Unobserved Preference Heterogeneity in Brand-Choice Models." *Marketing Science*, 25 (4): 322-35.
- Hsu, Cathy H.C, Oh, Haemoon., and Assaf, A. George (2011). "A Customer-Based Brand Equity Model for Upscale Hotels." *Journal of Travel Research*,

Hutchinson, J. Wesley, Wagner A. Kamakura, and John G. Lynch Jr. (2000). "Unobserved Heterogeneity as an Alternative Explanation for "Reversal" Effects in Behavioral Research." *Journal of Consumer Research*, 27 (3): 324-44.

Jedidi, Kamel, Jagpal, Harsharanjeet S., and DeSarbo, Wayne S. (1997). Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity. " *Marketing Science*, 1997 (16): 39-59.

Jedidi, Kamel and A. Ansari (2001), "Bayesian Structural Equation Models for Multilevel Data," in *New Developments and Techniques in Structural Equation Modeling*, Ed. George Marcoulides and Randall E. Schumacker, Laurence Earlbaum Associates, New Jersey.

Jurowski, Claudia, and Dogan Gursoy. (2004). "Distance Effects on Residents' Attitudes toward Tourism." *Annals of Tourism Research*, 31 (2): 296-312.

Kahn, Michael J., and Adrian E. Raftery. (1996). "Discharge Rates of Medicare Stroke Patients to Skilled Nursing Facilities: Bayesian Logistic Regression with Unobserved Heterogeneity." *Journal of the American Statistical Association*, 91 (433): 29-41.

Kass, Robert E., and Adrian E. Raftery. (1995). "Bayes Factors." *Journal of the American Statistical Association*, 90 (430): 773-95.

Kim, Kyungmi, Muzaffer Uysal, and M. Joseph Sirgy. (2013). "How Does Tourism in a Community Impact the Quality of Life of Community Residents?." *Tourism Management*, 36: 527-40.

Lee, Sik-Yum, and Xin-Yuan Song. (2003). "Bayesian Model Selection for Mixtures of Structural Equation Models with an Unknown Number of Components." *British Journal of Mathematical and Statistical Psychology*, 56 (1): 145-65.

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.

Li, Yong., and Wang, Hai.-Zhong. (2010). "Bayesian analysis for finite mixture in non-recursive non-linear structural equation models." *British Journal of Mathematical and Statistical Psychology*, 63 (2): 361-377.

Lubke, Gitta H., and Bengt Muthén. (2005). "Investigating Population Heterogeneity with Factor Mixture Models." *Psychological Methods*, 10 (1): 21-39.

Matzler, Kurt, Johann Füller, Birgit Renzl, Stephan Herting, and Sebastian Späth. (2008). "Customer Satisfaction with Alpine Ski Areas: The Moderating Effects of Personal, Situational, and Product Factors." *Journal of Travel Research*, 46 (4): 403-13.

Mazanec, Josef A., and Helmut Strasser. (2007). "Perceptions-Based Analysis of Tourism Products and Service Providers." *Journal of Travel Research*, 45 (4): 387-401.

Muthén, B., & Asparouhov, T. (2012). "Bayesian structural equation modeling: a more flexible representation of substantive theory." *Psychological methods*, 17(3): 313.

Nicolau, Juan L., and Francisco J. Mas. (2006). "The Influence of Distance and Prices on the Choice of Tourist Destinations: The Moderating Role of Motivations." *Tourism Management*, 27 (5): 982-96.

Nunkoo, R., Ramkissoon, H., & Gursoy, D. (2013). Use of Structural Equation Modeling in Tourism Research Past, Present, and Future. *Journal of Travel Research*, 52(6), 759-771.

Nylund, Karen L., Tihomir Asparouhov, and Bengt O. Muthén. (2007). "Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study." *Structural Equation Modeling*, 14 (4): 535-69.

Oh, Haemoon, and Cathy HC Hsu. (2014). "Assessing Equivalence of Hotel Brand Equity Measures in Cross-Cultural Contexts." *International Journal of Hospitality Management*, 36: 156-66.

Okazaki, Shintaro, and Morikazu Hirose. (2009). "Does Gender Affect Media Choice in Travel Information Search? On The Use of Mobile Internet." *Tourism Management*, 30 (6): 794-804.

Parasuraman, A, Zeithaml, Valarie A., and Berry, Leonard L. (1985). "A Conceptual Model of Service Quality and Its Implications for Future Research." *Journal of Marketing*, 49 (4): 41-50.

- Petrick, James F. (2004). "The Roles of Quality, Value, and Satisfaction in Predicting Cruise Passengers' Behavioral Intentions." *Journal of Travel Research*, 42 (4): 397-407.
- Preacher, Kristopher J., Michael J. Zyphur, and Zhen Zhang. (2010). "A General Multilevel SEM Framework for Assessing Multilevel Mediation." *Psychological Methods*, 15 (3): 209-33.
- Reisinger, Yvette., and Lindsay Turner. (1999). "Structural Equation Modeling with Lisrel: Application in Tourism." *Tourism Management*, 20 (1): 71-88.
- Rust, Ronald., and Verhoef, Peter. (2005). "Optimizing the Marketing Interventions Mix in Intermediate-Term CRM." *Marketing Science*, 24(3):477-489.
- Sarstedt, Marko, and Christian M. Ringle. (2010). "Treating Unobserved Heterogeneity in PLS Path Modeling: A Comparison of FIMIX-PLS with Different Data Analysis Strategies." *Journal of Applied Statistics*, 37 (8): 1299-318.
- Stepchenkova, Svetlana, Mills, Juline.E. (2010). "Destination Image: A Meta-Analysis of 2000–2007 Research." *Journal of Hospitality Marketing & Management*, 19 (6), 575-609.
- Temme, D., Williams, J., & Hildebrandt, L. (2002, January). Structural Equation Models for Finite Mixtures—Simulation Results and Empirical Applications. In *Compstat* (pp. 569-574). Physica-Verlag HD
- Wooldridge, Jeffrey M. (2005). "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics*, 20 (1): 39-54.
- Weed, Mike. (2006). "Sports Tourism Research 2000–2004: A Systematic Review of Knowledge and a Meta-Evaluation of Methods." *Journal of Sport & Tourism*, 11(1):5-30.
- Wu, Paul, Gary Yeong-Yuh Yeh, and Chieh-Ru Hsiao. (2011). "The Effect of Store Image and Service Quality on Brand Image and Purchase Intention for Private Label Brands." *Australasian Marketing Journal*, 19 (1): 30-9.
- Yoon, Yooshik, and Muzaffer Uysal. (2005). "An Examination of the Effects of Motivation and Satisfaction on Destination Loyalty: A Structural Model." *Tourism Management*, 26 (1): 45-56.

Yuan, Ke-Hai, and Peter M. Bentler. (2010). "Two Simple Approximations to the Distributions of Quadratic Forms." *British Journal of Mathematical and Statistical Psychology*, 63 (2): 273-91.

Zhang, Heng, and Siu Lai Lei. (2012). "A Structural Model of Residents' Intention to Participate in Ecotourism: The Case of a Wetland Community." *Tourism Management*, 33 (4): 916-25.

Zyphur, Michael J., Sankalp Chaturvedi, and Richard D. Arvey. (2008). "Job Performance Over Time Is a Function of Latent Trajectories and Previous Performance." *Journal of Applied Psychology*, 93 (1): 217-24.

Figure 1. A Lodging Brand Equity Model

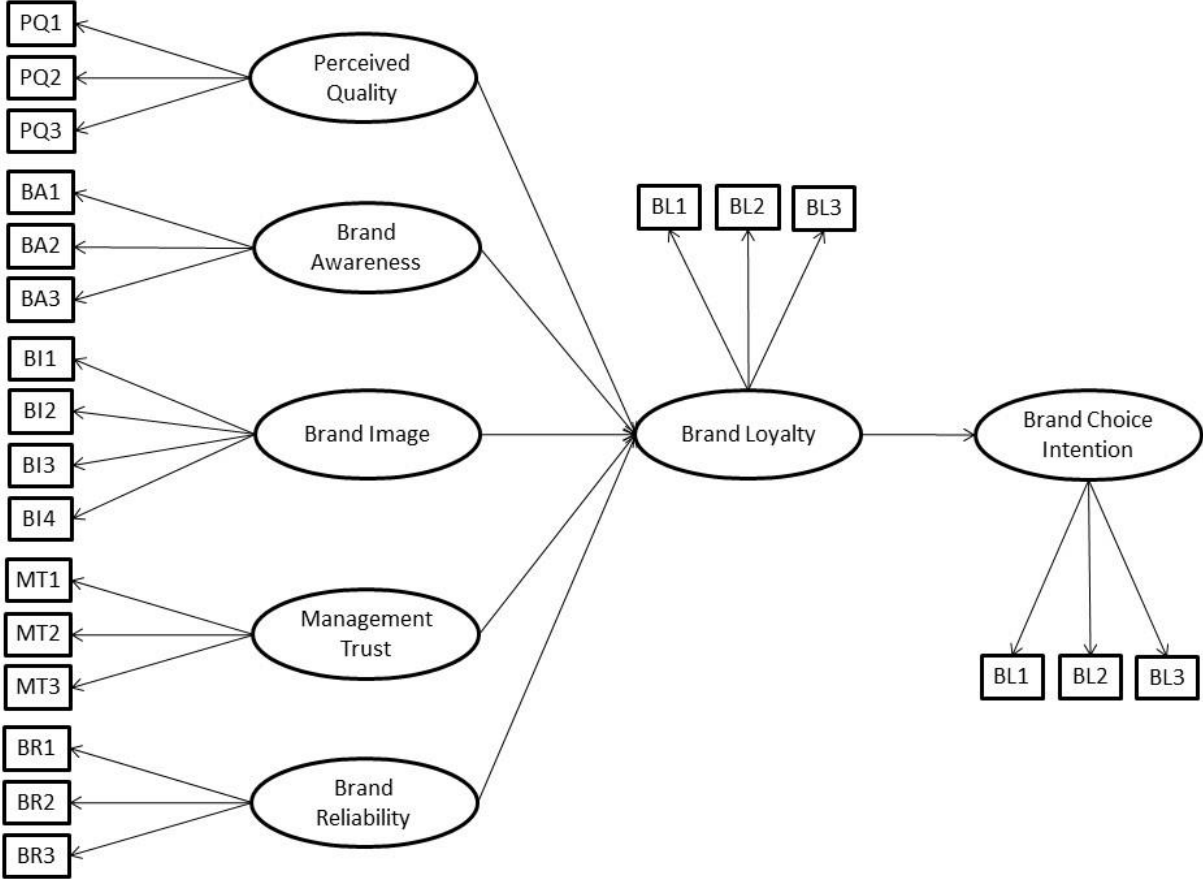


Table 1. Data Characteristics

Construct and Measurement Item	Mean	SD	Skewness	Kurtosis
Brand choice intention				
Even if other competing brands are not different from XYZ in any way, it seems smarter to choose an XYZ hotel.	5.09	1.24	-0.54	0.33
An XYZ hotel is always a superior choice to its rival hotels.	4.99	1.25	-0.36	0.04
It makes sense to choose XYZ instead of any other hotel brand, even if they are the same.	4.90	1.32	-0.38	0.10
<i>XYZ is my favorite brand of all competing hotel brands.</i>				
Brand loyalty				
I will choose XYZ hotels over and over again without hesitation.	4.90	1.42	-0.52	-0.11
I feel good and positive when I think about staying at an XYZ hotel.	5.20	1.25	-0.58	0.20
Thinking about the XYZ hotel makes me feel pleasant.	5.17	1.29	-0.56	0.18
<i>I would not choose other hotel brands if a XYZ hotel is available when I need accommodation.</i>				
Brand quality				
XYZ is of high quality.	5.55	1.07	-0.58	0.33
XYZ sets quality standards other hotels should follow.	5.32	1.17	-0.57	0.32
I consider XYZ's quality to be of the highest standard.	5.17	1.25	-0.75	0.66
<i>XYZ is regarded as a leader in quality.</i>				
Brand awareness				
I know what the XYZ symbol or logo looks like.	5.24	1.52	-0.86	0.32
I know what an XYZ hotel looks like.	5.17	1.35	-0.75	0.34
I can easily recognize XYZ hotels among other competing hotels.	5.17	1.41	-0.72	0.17
<i>XYZ is unique and different from other hotel brands.</i>				
Brand image				
The XYZ brand is prestigious.	5.24	1.20	-0.63	0.55
The XYZ hotels tend to attract sophisticated people as guests.	5.18	1.22	-0.61	0.42
Staying at an XYZ hotel makes me feel special.	5.05	1.28	-0.60	0.30
I have a clear image of the type of people who would stay at an XYZ hotel.	4.74	1.38	-0.39	-0.21
Management trust				
I trust the XYZ management.	5.44	1.15	-0.60	0.35
The XYZ management knows how to do the hotel business.	5.42	1.15	-0.54	0.18
The XYZ implements good management practices other hotels can learn.	5.29	1.20	-0.51	0.15
<i>XYZ management is sophisticated enough not to ignore customer problems.</i>				
Brand reliability				
The XYZ will meet my expectations every time.	5.25	1.18	-0.52	0.20
My experience with XYZ will be consistent every time I stay.	5.27	1.16	-0.45	0.07
The XYZ will not disappoint me every time.	5.13	1.22	-0.54	0.33
<i>I feel safe and secure when staying at XYZ hotel.</i>				

*Items in italics were dropped from the estimation.

* This table is adapted partly from Hsu et al. (2013).

Figure 2. Homogenous SEM

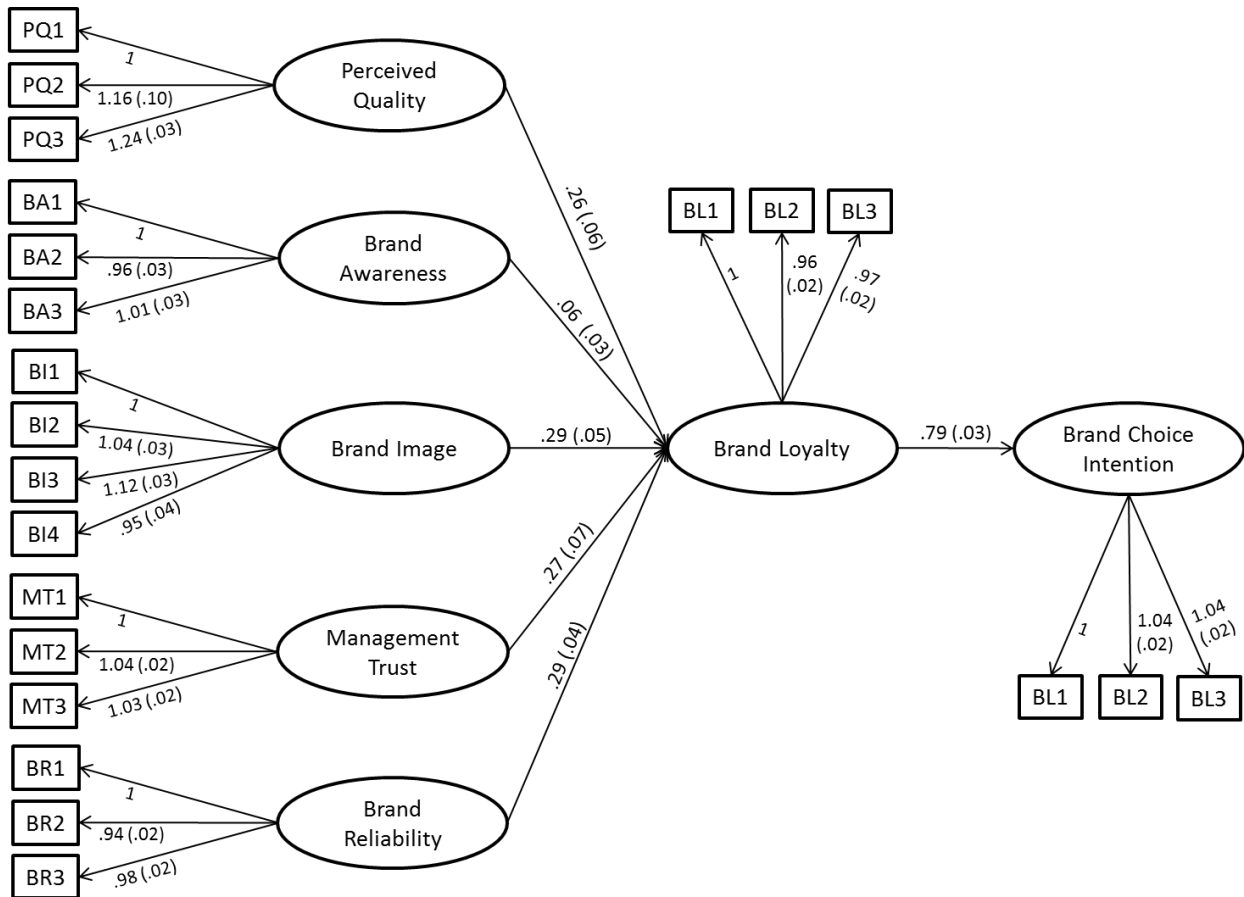


Table 2. Deviance Information Criterion (DIC) for Model Comparison

Model	DIC
Homogeneous SEM	60396
The FMSEM with two groups	32334
The FMSEM with three groups	35339

Figure 3. The FMSEM: Group 1

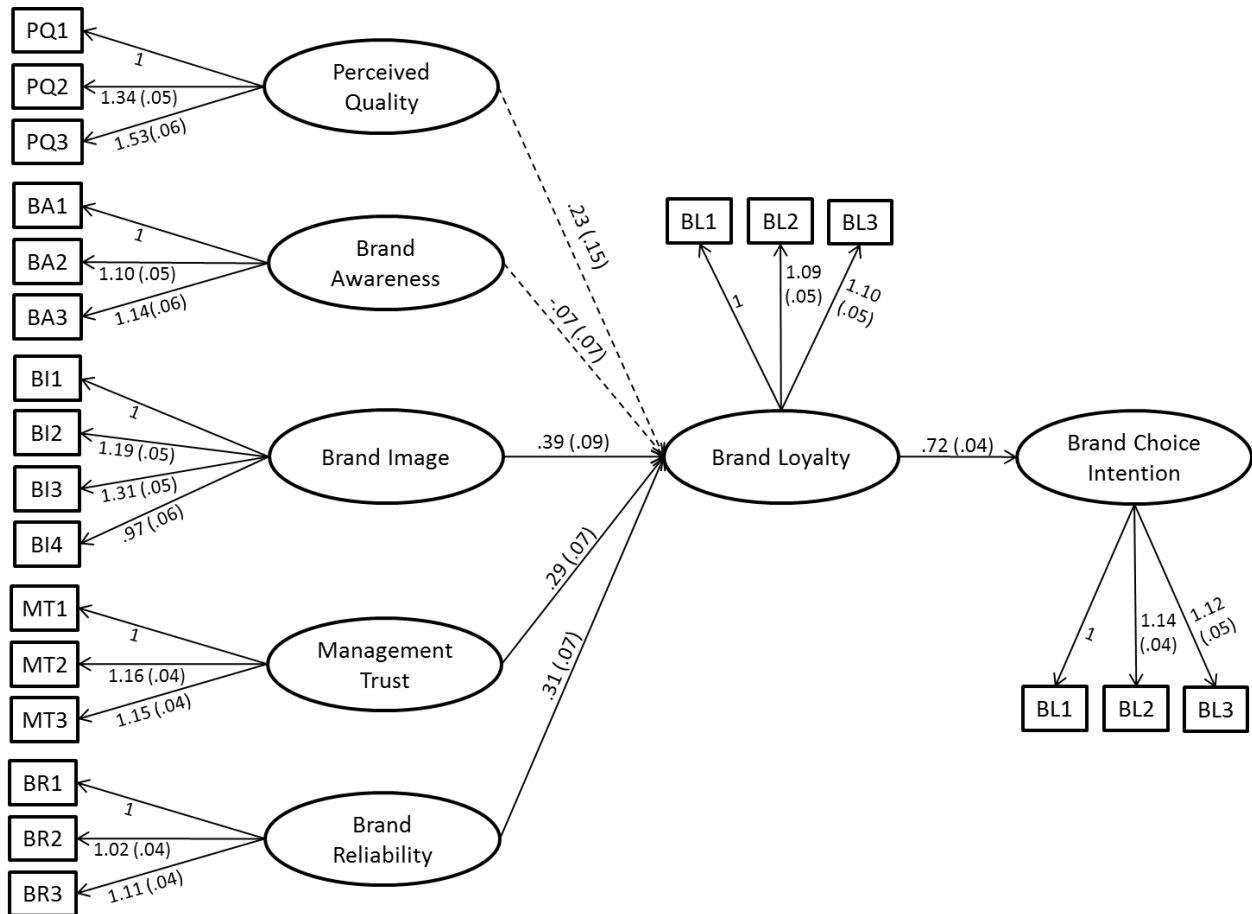


Figure 4. The FMSEM: Group 2

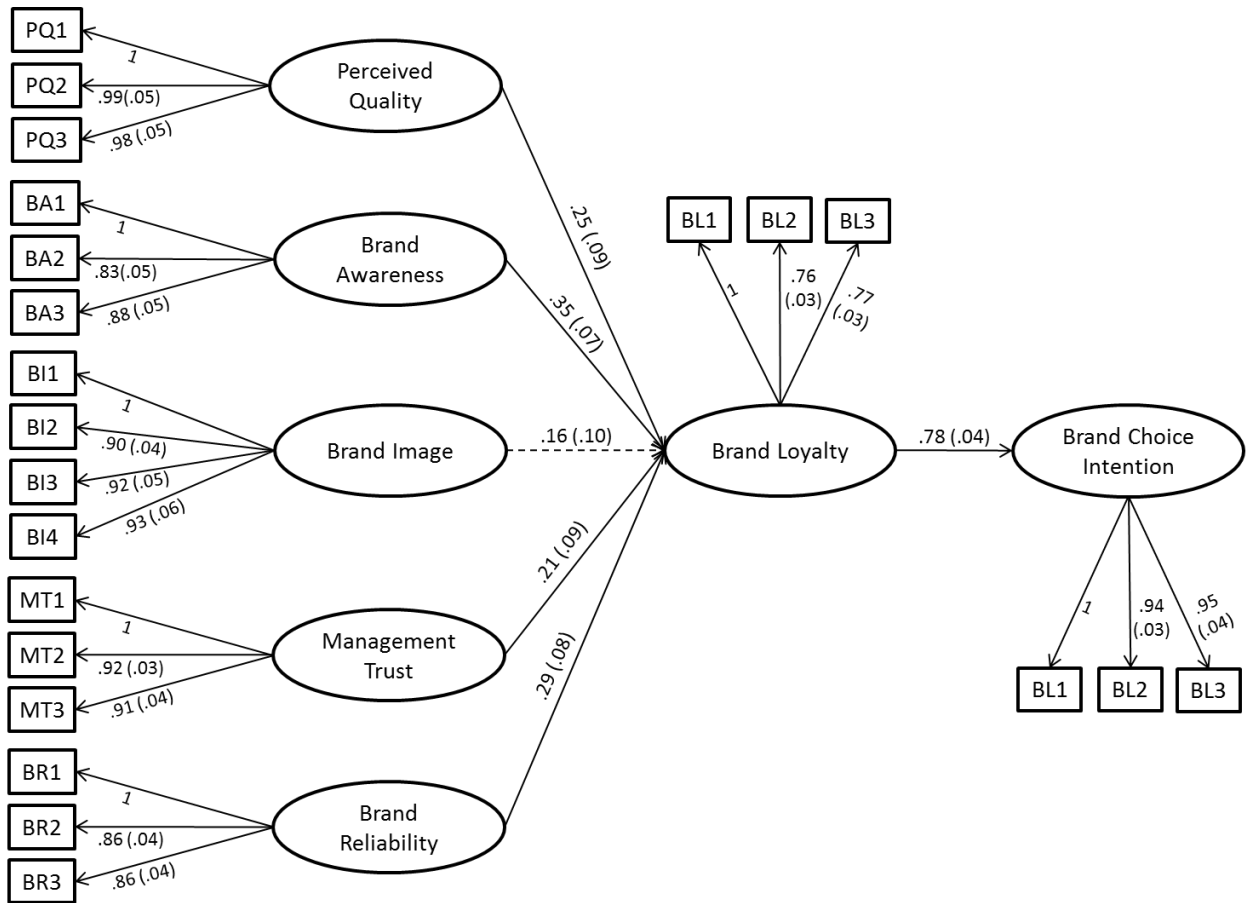


Table 3. Covariates of Unobserved Heterogeneity

Variable	Estimate	Standard deviation
Perceived Quality	0.45*	0.27
Brand Awareness	0.23	0.27
Brand Image	-1.18**	0.30
Management Trust	-0.23	0.18
Brand Reliability	0.19	0.32

Appendix 2: Monte Carlo Experiment

As mentioned to further validate the performance of our model we conducted a Monte Carlo experiment using the same data on the regressors from our application. In this simulation experiment we consider as true values of the parameters the posterior means obtained from the Bayesian estimation of our models. Our interest focuses on whether the Bayesian techniques can recover the true number of groups with $N=1,346$ and also $N=2,700$, $N=800$, $N=400$ and $N=100$ using the DIC criterion. We use 10,000 Monte Carlo replications and we use 5,000 preliminary draws followed by another 10,000 to obtain posterior statistics. The true data generating process is taken to be a homogeneous SEM and an SEM with two and three groups. The Monte Carlo experiment consists of estimating the three models by MCMC and computing the DIC as well as the mean squared error (MSE) of the parameters (relative to the “true” parameters that we set for the Monte Carlo experiment). For the methods to perform well it must be the case that (i) MSEs are “small” and fall as the sample size increases and (ii) the number of times the true model is selected using the DIC is “large” and increases with the sample size. In Table A.1 we report the number of Monte Carlo samples for which the true process is selected using the DIC statistic. Reported also are average mean squared errors of all estimated parameters. It turns out that even in samples as small as 100, the Bayesian approach can be used profitably to select the true model with probability 72.17% for homogeneous models, 65.23% for models with 2 unobserved groups and 58.23% for models with three groups. This probability goes up to 77.23% in samples of size 400 and 88.97% when the sample size is 800 –closer to what we actually observe in the typical case. The mean squared errors (MSEs) of parameters are small and drop roughly as the squared root of N . Clearly, the Bayesian approach is quite useful in the sense that in samples of size 800 or higher, the DIC’s discriminatory power is substantial (88.97%) in models with three groups. Using the results of the Monte Carlo simulation we can be fairly confident that the homogenous model is not true and that our model is revealing the true number of groups. Our evidence is reinforced by the fact that our Monte Carlo experiment uses the same data as the data that we use in our empirical application and the estimated parameters.

Table A1. Monte Carlo results

	True Data Generating Process					
	<i>Homogenous SEM</i>		<i>Two Groups</i>		<i>Three Groups</i>	
N=100	72.17% ⁽¹⁾	0.103 ⁽²⁾	65.23% ⁽¹⁾	0.253	58.23% ⁽¹⁾	0.355
N=400	87.21% ⁽¹⁾	0.049	81.35% ⁽¹⁾	0.107	77.23% ⁽¹⁾	0.145
N=800	93.21% ⁽¹⁾	0.035	91.10% ⁽¹⁾	0.088	88.97% ⁽¹⁾	0.091
N=1,346 ⁽³⁾	100% ⁽¹⁾	0.027	98.17% ⁽¹⁾	0.073	95.21% ⁽¹⁾	0.081
N=2,700	100% ⁽¹⁾	0.019	100% ⁽¹⁾	0.063	99.77% ⁽¹⁾	0.070

Notes: (1) represents the percentage of Monte Carlo samples for which the true process is selected using the DIC statistic. (2) represents average mean squared errors (MSEs) of all estimated parameters, collectively. (3) $N=1,346$ corresponds to the sample size in our application and $N=2,700$ to, roughly, twice this sample size.