Structural Equation Modeling in Tourism Demand Forecasting: 
A Critical Review

Guy Assaker*, Vincenzo Esposito Vinzi**
& Peter O’Connor**

*Lebanese American University/Lebanon
**ESSEC Business School, Avenue Bernard Hirsch/France

Received (in revised form): August 2010

Abstract
This paper reviews prior applications of structural equation modeling to tourism demand forecasting in seven prominent tourism and service journals over a twelve-year period, 1998-2009. Having identified the applications over time at both the individual and aggregate demand levels, this paper discusses essential methodological issues related to structural equation modeling as it applies to tourism demand modeling. It then assesses the quality of previous applications in terms of six methodological issues: (1) data characteristics and sample size; (2) model identification; (3) overall model fit; (4) reliability and validity; (5) model re-specification; and (6) reporting. Based on this analysis, the discussion identifies several methodological problems in the use of SEM in tourism demand modeling and suggests specific avenues for improvement.

Keywords: Tourist demand, tourist behavior intention, structural equation modeling, SEM.

Introduction

Tourism has become a major force in today’s world economy. The travel and tourism industry is one of the largest single employers and, in many countries is the largest contributor to the services export sector, significantly affecting balance of payments (Papatheodorou, 1999). According to the World Travel and Tourism Council, more than 255 million people (or 11.1% of total jobs worldwide) are employed in tourism related activities (WTTC, 2008). Furthermore, since the 1940s, international travel has transformed from being an activity solely for the rich into an activity for the population at large (Ioannides & Debbage, 1997). According to the World Tourism Organisation, almost 922 million people travelled abroad for tourism in 2008, generating more than US$944 billion in revenues. Considered as a whole, these expenditures accounted for nearly two percent of the world’s gross national product (UNWTO, 2009).

Accompanying this growth in international tourism is increased interest in tourism research (Song & Li, 2008). The topic of demand modeling and forecasting has attracted the attention of both academics and practitioners as an important area of inquiry within tourism (see e.g., Archer, 1987; Van doorn, 1982; Witt & Witt, 1992; Witt & Song, 2000; Li, Song, & Witt, 2005). Accurately forecasting demand is essential for efficient planning...
by airlines, shipping companies, railways, hoteliers, tour operators, food and catering establishments, and other sectors connected with tourism (Crouch, 1995). Li et al., (2005) argue that the need to forecast accurately is especially acute because the tourism product is perishable. Unfilled airline seats and unused hotel rooms cannot be stockpiled; thus, demand must be anticipated and manipulated. Forecasts are also of great interest to governments and national tourist organizations to keep pace with the rapid flow of tourists. Witt and Song (2000) argued that because both the public and private sector must plan for tourism demand to invest in appropriate infrastructure, understanding the demand for tourist arrivals in terms of both volumes and determinants can have important economic consequences.

To understand the relationship between tourist arrivals, their determining factors, and subsequently estimate future tourist volume, most studies of tourism demand have traditionally used econometric models (also known as causative models) (Li et al., 2005). These have evolved around regression models to estimate international tourist flows between countries (Morley, 1998; Witt & Wilson, 1994). Such models have been predominantly single equation models that explain demand at the aggregate level. More recently, however, scholars have begun using more advanced research methodologies for econometrics modeling (Kennedy, 2003). In addition to Structural Equation Modeling, modern econometric methods include the error correction model; time varying parameter model; and autoregressive distributed lag model. Each of these has been used as an alternative method to traditional econometrics in general demand forecasting (Armstrong, 2001). With the exception of SEM, each has been widely adopted and applied in tourism research. Witt and Witt (1995), Li et al. (2005), and Song and Li (2008) explained these models and their applications to tourism demand forecasting in detail.

SEM, however, has had to date limited applications in tourism demand modeling compared to other methods (Li et al., 2005). This may be due to the complexity of the SEM method and the lack—until recently—of user-friendly computer manuals (Reisinger & Turner, 1999). However unlike traditional multivariate regression models, SEM uses simultaneous equation models in which variables (both observed and latent) may influence one another reciprocally. This makes SEM a very suitable method for analyzing tourism demand at both the individual (motivations and behavior) and aggregate (where combined individual determinants explain patterns of tourism demand) levels, as it allows representations of causal relationships among a multitude of variables related to personal determinants and the supply and demand environments, thereby explaining tourism demand at the destination stage (Song & Li, 2008).

The purpose of this paper is two-fold. First, it documents how SEM has been applied from a technical perspective in studies published in seven major tourism and service industry journals. These journals are Journal of Travel Research, Tourism Management, Tourism Economics, Annals of Tourism Research, Journal of Hospitality and Tourism Research, Journal of Service Research, and Journal of Services Research. The paper then critically evaluates how SEM has been used in published papers and provides guidance for future users on how to employ the methodology correctly. The paper proceeds as follows: first, the paper emphasizes how SEM can suitably model tourist demand under various perspectives at both the aggregate and individual levels. Second, the paper explains the concept of SEM modeling. It then reviews how SEM has been applied to tourism
modeling in each of the published papers identified. The paper then evaluates the methodological quality of applications by assessing how they conform to formal statistical assumptions required for the valid use of these techniques while identifying problem areas and suggesting avenues for improvement. Finally, the paper concludes by summarizing the findings and results and providing a checklist of technical issues to consider when using SEM methodology in tourism demand modeling.

Why Use SEM to Forecast Tourism Demand?

At the individual level, SEM is particularly appropriate for tourism research because the factors influencing tourism demand are linked to personal determinants of consumer behavior (Smith, 1994). For example, differences in attitudes, perceptions, travel motivators, and images of destinations are critical in the travel decision process (Sirakaya & Woodside, 2005). Indeed, given that personal determinants are latent, the SEM model can be best used to represent the observed dimensions of an unobserved structure.

SEM is also appropriate at the aggregate level because it can represent causal relationships among a variety of variables related to the supply and demand environments. Prior to SEM, these variables were difficult to incorporate into a single economic model. For example, Cooper and Wahab (2001) argued that given the difficulties of relating demand volume to many variables simultaneously, researchers have attempted to isolate the most influential group of variables and subsequently relate the demand volume to changes in these key variables. Zhang & Jensen (2007) argued that economic determinants, which are largely related to the dimension of generating tourism demand, were the easiest to measure and thus have been commonly employed in tourism demand studies. Consequently, SEM can operationalize these held-out determinants and investigate their causal relationships to determine tourist inflows to a destination under study.

Like any advanced statistical tool, however, SEM requires several complicated choices, which users must understand in order to apply it correctly. By using SEM correctly, researchers can avoid missteps that could compromise the validity of results; are restrained from inferring incorrect conclusions; and can develop accurate knowledge about causal relationships among variables (Fan, Thompson, & Wang, 1999). The complications in applying SEM have been highlighted in many other fields, for example organizational behavior (Brannick, 1995), strategic management (Shook, Ketchen, Hult, & Kacmar, 2004), management information systems (Chin, 1998), marketing (Baumgartner & Homburg, 1996), and logistics (Steiger, 2001). To date, however, no systematic assessment has been made of how SEM is used within tourism demand studies.

Given the limited application of SEM in tourism demand modeling to date, the appropriateness of such methods in tourism demand studies at both individual and aggregate levels, and the difficulties involved in correctly applying this type of statistical method, a comprehensive examination of SEM application as it applies to tourism modeling is timely and warranted. The next section reviews the general objectives and advantages of SEM.
Objectives And Advantages of SEM

SEM’s primary aim is to depict the pattern of a series of inter-related dependent relationships simultaneously among a set of latent constructs, each measured by one or more manifest variables (Reisinger & Tuner, 1999). Using SEM, latent constructs are labeled exogenous (independent) and endogenous (dependent). SEM includes one or more linear regression equations describing how the endogenous constructs depend upon exogenous and other endogenous constructs (Byrne, 2001). Their coefficients are called path coefficients or regression weights. Based on this understanding, SEM is similar to combining factor analysis and multiple regressions.

SEM differs from factor analysis modeling, however. In factor analysis (e.g., principal components analysis), links between the observed and latent variables (factors) are unknown or uncertain. As such, the analysis proceeds as exploratory to identify the minimum number of factors that account for covariations among the observed variables. In contrast, when using SEM, the relationships between the observed variables and underlying factors are postulated a priori; and confirmatory factor analysis (CFA) is then used to test the hypothesized structure statistically (Schumacker & Lomax, 2004). By specifying the pattern of inter-variable relations a priori, SEM lends itself well to data analysis for inferential purposes. SEM differs in this way from other factor analysis procedures, which are descriptive by nature, making hypothesis testing difficult or impossible. Because confirmatory factor analysis focuses only on the link between factors and their measured variables, within the SEM framework, it represents what has been termed a measurement model (Ullman & Bentler, 2004).

Important differences also exist between traditional regression models and SEM modeling. Traditional regression models can neither assess nor correct measurement error because alternative methods rooted in regression or the general linear model assume that errors in the explanatory (independent) variables vanish. SEM, however, provides explicit estimates of the error variance parameter. Moreover, although data analyses using regression methods are based on observed measurements only, those using SEM procedures can incorporate both unobserved and observed variables. Additionally, SEM is particularly useful because one dependent variable can simultaneously cause another. In this case, SEM is a powerful method for tracing direct and indirect effects and effectively dealing with multicollinearity, thus representing another benefit of SEM over traditional regression models (Byrne, 2001).

Finally, SEM techniques can represent and analyze models with feedback loops. Many real world causal processes are based on cycles of mutual influence; that is, feedback. The presence of a feedback loop in SEM automatically makes it nonrecursive. As described in the methodological issues section of this paper, however, it is often more difficult to analyze nonrecursive models because of identification and results stability issues (Kline, 2004).

Moreover, all aspects of SEM modeling (measurement and regression models) must be directed by theory, which is critical for developing and modifying models. A clear misuse of SEM can occur when data are simply fitted to a suitable SEM and results are used to draw causal arrows in models or resolve causal ambiguities. Thus, theoretical insight and judgment are of paramount importance in SEM. In addition, SEM models are
commonly estimated using covariance (structure) analysis. This means that the model parameters are determined such that the variances and covariances of the variables implied by the model system are as close as possible to the observed variances and covariances of the sample. As such, SEM is also referred to as covariance structure analysis, covariance structure modeling, and analysis of covariance structures.

Although these alternate references accurately indicate that SEM focuses on analyzing covariance, SEM can also analyze a model’s mean structure. If only covariances are analyzed, then all observed variables are mean-deviated (centered) so that latent variables must have a mean of zero. Sometimes this loss of the means is too restrictive. When the focus extends to analyzing mean structures, the means and intercepts also become central parameters in the model. The most common type of SEM uses a mean structure to estimate the parameters of growth, referred to as a latent growth curve SEM model. Another scenario in which calculating the means of latent parameters is useful is when the same model is tested across samples to check for similarities/invariances in parameters; namely, multi-group analysis or multi-sample SEM analysis.

An alternative method of modeling the relationship among latent variables is partial least squares (PLS) path modeling, which generates path coefficients for a SEM-type model. PLS path modeling is “soft modeling” because it is not a covariance-based modeling technique. Instead, it uses heuristics and makes relaxed assumptions about data and model specification. These two later points are considered limitations for using the covariance-based SEM technique. Henseler, Ringle, & Sinkovics (2009) argued, however, that with small sample sizes; when the model analyzed is complex with many latent and manifest variables; or in the presence of both reflective and formative measurement models, PLS path modeling is methodologically preferred. PLS applications, however, are very limited in the tourism literature, with only one paper found using a tourism application (Arteaga, Gallarza, & Gill, 2010) while collecting data for the current study.

To summarize, SEM in comparison to the so-called first generation multivariate procedures—in particular, factorial analysis and traditional multiple regression models—offers specific advantages that have been demonstrated in many tourism demand-related applications. These are reviewed in the next section.

**Methodology**

For this study, we selected five journals as the most representative of research in the general field of tourism to examine how SEM has been applied to tourism demand modeling. These five are *Annals of Tourism Research, Journal of Hospitality and Tourism Research, Journal of Travel Research, Tourism Economics,* and *Tourism Management*. The papers were chosen based on the classification made by Murphy and Law (2008), who ranked 50 tourism journals based on the Google Scholar’s average annual citations from the first year of publishing until July 2007. This period includes the approximate period examined in this study. This classification method is considered a good proxy for the traditional Journal Citation Report (JCR) method undertaken by the Institute for Scientific Information (ISI). The JCR could not be used, because only two journals in the tourism discipline (Annals of Tourism Research and Tourism Management) are included so far in the ISI database (Law & Van der Veen, 2008). The chosen journals were the top five
ranked in Murphy and Law’s (2008) classification table after excluding specific-industry-based journals.

Because tourism is a fundamental element of the global service industry, the *Journal of Services Research* and *Journal of Service Research* were also examined for potential SEM applications in tourism demand forecasting. The choice of these two journals was based on a two-step selection process. First, we identified service journals considered in the ISI database. Then, we used the Google Scholar *advanced search* option, where we listed the name of the journal in the *publication field* and “tourism” as the *keyword* to search for *anywhere in the article* for the *studied period* of 1998 to 2009. This process identified the number of tourism applications in each of the service journals available in the ISI database. The two journals noted were deemed the most relevant as they published the larger number of tourism papers for the considered period.

The 12-year period of study between 1998 and 2009 was considered appropriate because SEM is a recent application used generally in research and specifically in tourism (Li et al., 2005). Papers using exploratory factor analysis models and other path models estimated with traditional regression methods (i.e., Ordinary Least Square) were excluded from the sample. Moreover, as stated, although PLS is somewhat similar to SEM, it is not a covariance-based modeling technique; therefore, its application differs significantly. Thus, the current research disregarded studies using PLS. (Indeed, when reviewing journals, only one study using PLS in demand modeling was identified.).

Twenty-one papers satisfied the selection criteria. The identified SEM applications included confirmatory measurement studies (1 study, 4%); single-indicator structural path studies (4 studies, 19%); full structural model studies (14 studies, 67%); and multiple indicators multiple independent causes (MIMIC)\(^1\) studies (2 studies, 10%). Table 1 lists the Journals, authors, and type of study. None of the reviewed papers used SEM models with feedback loops (non-recursive). Moreover, the reviewed papers solely used covariance-based structure; subsequently, none of the reviewed studies analyzed the intercept and mean of the constructs in the context of multi-group cross sectional analysis or latent growth analysis (to examine parameters of change in studied variables). These findings suggest potential further extensions of tourism demand modeling using SEM.

As a final characterization, seventeen studies (81%) used the individual level of analysis compared to only four studies (19%) at the aggregate level of analysis. The discrepancy in the number of papers covering individual versus aggregate demand analysis suggests that opportunities remain to apply SEM at the macro demand forecasting level.

---

\(^1\) MIMIC studies refer to models in which some constructs have both reflective (causal relationships from constructs to variables) and formative (causal relationships go from variables to constructs) indicators.
Table 1: Tourism Demand Articles Using SEM

<table>
<thead>
<tr>
<th>Annals of Tourism Research</th>
<th>Tourism Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baker and Crompton, 2001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Bigné, Sanchez, and Sanchez, 2001&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Lehto, O’Leary, and Morisson, 2004&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Gallarza and Saura, 2004&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Um, Chon, and Ro, 2006&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Lam and Hsu, 2005&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Hyde, 2008&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yoon and Uysal, 2005&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Chen and Tsai, 2007&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Journal of Hospitality and Tourism Research</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsu, 2004&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Jang and Feng, 2007&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ryu and Jang, 2006&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Chi and Qu, 2008&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Jang, Bai, Hu, and Wu, 2009&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Zakbar, Bencic, and Dmitrovic, 2009&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Journal of Travel Research</th>
<th>Journal of Service Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petrick, Morais, and Norman, 2001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Lacey, Suh, and Morgan, 2007&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mazanec, Wober, and Zins, 2007&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tourism Economics</th>
<th>Journal of Services Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuner, Reisener, and Witt, 1999&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Chathoth, 2001&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Tuner and Witt, 2001&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> SEM models with a total of 14 studies (67%)
<sup>b</sup> MIMIC models with a total of 2 studies (10%)
<sup>c</sup> Path models with a total of 4 studies (19%)
<sup>d</sup> CFA models with a total of 1 study (4%)

Articles in bold refer to SEM studies at the tourist aggregate demand level.

Methodological Issues In Applying SEM

This section examines the methodological use of SEM in tourism demand papers, focusing on six critical issues as summarized by Shook et al. (2004) (1) data characteristics and sample size; (2) model identification; (3) overall model fit; (4) reliability and validity; (5) model respecification; and (6) reporting. Best practices (relying on statistical tests or rules of thumb) were derived from standard literature on SEM methodology (e.g., Bollen, 1989; Fornell & Larcker, 1981; Joreskog & Sorbom, 1996) to evaluate the researchers’ decisions. The aim was to identify tendencies in the methodological use of SEM in tourism concerning each of these issues and also to benchmark results to best practices in the SEM literature. Suggestions to improve employing the methodology more appropriately are also discussed.
Data Characteristics and Sample Size

**Missing data:** Most SEM computer programs do not yield results in the presence of missing data unless intercept and mean are evaluated concurrently with the covariance structure (Kaplan, 1990). For example, AMOS automatically uses the full information maximum likelihood (FIML) method to replace missing values in the dataset when mean structure analysis is considered (Kline, 2004). Thus, in order for results to be valid, missing values must be dealt with properly.

Most methods to deal with incomplete observations assume that the cases are missing completely at random (MCAR; Vriens & Metlon, 2002). Such methods have recently become available in SEM programs such as LISREL, EQS, and AMOS (Ullman & Bentler, 2004). Listwise deletion refers to a case with missing values that is ignored in all calculations, whereas pairwise means it is ignored only for calculations involving that variable. The pairwise method can result in correlations or covariances falling outside the range of the possible (Kline, 1998), which can lead to singular (non-positive definite) covariance matrices. This prevents such math operations as inverting the matrix because division by zero will occur. This problem does not occur with listwise deletion. Thus, given that SEM uses covariance matrices as input, listwise deletion is recommended when the sample is large and the number of cases to be dropped is small. An effective rule of thumb is to use listwise deletion when it will lead to eliminating 5% of the sample or less (Kline, 2004).

When listwise deletion cannot be used, some form of data imputation is recommended. In mean imputation, the mean of the variable is substituted. This approach is not recommended because it shrinks the variances of the variables involved (Kline, 2004). Regression imputation predicts the missing value based on other non-missing variables. In pattern matching, missing data are replaced by the response to that variable based on a case whose values match the given case in all other variables. Finally, maximum likelihood imputation replaces missing values with estimates that maximize the probability of observing what has, in fact, been observed and has been suggested to yield the least bias (Oilinsky et al., 2003).

Among the 21 papers studied, the issue of missing data was addressed in three (14%) papers (see Table 2), none of which communicated on the level of missing data. Moreover, only one of the papers tested the missing data pattern and demonstrated it to be missing at random. Mean imputation methods, maximum likelihood imputation, and listwise were respectively used in the three papers. These results suggest that tourism researchers often neglect to inform readers how missing data are handled in SEM analysis. Future researchers should communicate the percentage of missing cases and the pattern of the missing data, because such information is critical for choosing and justifying the appropriate remedy. Moreover, researchers should be aware of the limitations of the various remedies employed and their potential impact on the findings.
Table 2: Statistics Concerning Data Characteristics and Sample Size Issues in Applying SEM to Tourism Demand Modeling

<table>
<thead>
<tr>
<th>Data characteristics</th>
<th>Number of Papers (N = 21)</th>
<th>Frequency (N = 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collected from primary sources</td>
<td>17</td>
<td>81%</td>
</tr>
<tr>
<td>Data collected from secondary sources</td>
<td>4</td>
<td>19%</td>
</tr>
<tr>
<td>Normality</td>
<td>4 (Yes)</td>
<td>19%</td>
</tr>
<tr>
<td>Missing data</td>
<td>3 (Yes)</td>
<td>14%</td>
</tr>
<tr>
<td>Covariance or correlation matrix</td>
<td>Cov = 4</td>
<td>19%</td>
</tr>
<tr>
<td>Average # of estimated parameters</td>
<td>16</td>
<td>%</td>
</tr>
<tr>
<td>Average sample size</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td>Average ratio of sample size/parameters</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Min. ratio of sample size/parameters</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Max. ratio of sample size/parameters</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

**Normality and linearity:** Data must also meet the assumed distribution of the estimation approach used. Maximum Likelihood Estimation (MLE) is the dominant approach for estimating SEM; it assumes that indicator variables have multivariate normal distributions. Simulation studies (Kline, 2004) have suggested that situations of severe non-normality of data may lead to inflated goodness-of-fit statistics and underestimated standard errors. This would indicate that regression paths and factor/error covariances were found to be more statistically significant than they should be which may mislead research progress by providing inaccurate findings (MacCallum, Roznowski, & Necowitz, 1992).

Despite these concerns, only two (10%) of reviewed studies noted whether the sample was normally distributed (see Table 2). Indeed, normality is routinely presumed in the case of primary data involving Likert scales when the number of Likert categories is four or higher (Joreskog & Soborm, 1988). Among the 21 papers reviewed, 17 (81%) used primary data collection with a Likert scale of four categories or more, while four papers used a secondary source of data (see .2). The two papers in which normality was discussed are among the four papers in which data were collected from secondary sources. The authors further took corrective actions (e.g., transforming the data using log and square root alterations) to achieve normality for the data prior to fitting the model.

Data transformation, however, is not without problems. Researchers can use many types of transformations, and it may be necessary to try several different transformations before finding one that works for a particular distribution. Moreover, some distributions can be so severely non-normal that, essentially, no transformation will work. According to Satorra (2001), to the extent that a researcher developed a strong theoretical foundation and believed in the original specification, transforming the data can provide an incorrect...
specification. Thus, researchers must be aware of the inherent detriment to transforming data (Shook et al., 2004). An alternative to transforming data is an estimation strategy, such as generalized least squares, which does not assume multivariate normality (Hayduk, 1987). An inherent limitation of such an alternative method when normality of data is not met, however, is that large samples (n > 2,500) are required for these methods to work well (Byrne, 2001).

Although just two of the 21 studies checked for normality, all 21 articles used MLE. None of the studies used alternative estimation techniques, including the four papers in which secondary data collection methods were applied and normality should, in general, be checked.

Furthermore, SEM estimates the parameters that best reproduce the sample covariance matrix, and the covariance matrix subsequently assumes linearity between variables. SEM further assumes linear relationships between indicator and latent variables and between latent variables themselves. This violates the linearity assumption, meaning that estimates of model fit and standard error become biased (i.e., not robust). Because non-linear modeling usually involves violating the assumption of multivariate normality (Hair, Black, Babin, Anderson, & Tatham, 2010), some researchers (e.g., Kline, 2004) have advocated that transformations of the measured variable (exponential, logarithmic, or other non-linear transformations) to achieve normality also correct for linearity.

In the 21 papers examined, none addressed the issue of linearity. These results indicate that by routinely choosing the most commonly used estimation technique (i.e., MLE) without a priori support of its use, not enough attention is being paid to data screening prior to estimating the model. It should be noted, however, that in their assessment of SEM in strategic management research, Shook et al. (2004) argued that data screening is probably not described in most studies due to space constraints. Future research should consider multivariate normality more seriously and include a summary measure describing the extent to which the data are normally distributed. This could be included in the methods or results section, especially when a secondary source of data is used. Researchers can improve their future applications of SEM by testing the data for linearity prior to running the SEM analysis. The Bartlett test and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy can be used to examine linearity on the variables’ overall basis (Hair, Black, Babin, Anderson, & Tatham, 2010).

**Type of data:** In addition to data screening, it is important to determine the type of data being examined (e.g., categorical, ordinal, or numerical), and which measure of association should subsequently be used in the analysis. Exogenous variables in SEM may be dichotomies or dummy variables, but this is not the case for endogenous variables. Unless special approaches are taken, categorical dummy variables may not be used as endogenous variables (Long, 1997). Endogenous variables should be continuous with normally distributed residuals, because using an ordinal or dichotomous measurement to represent an underlying continuous variable leads to attenuation of the coefficients in the correlation matrix that SEM uses (Lee, 2007).

For this reason, if ordinal data are to be modeled, they should be modeled as ordered-categorical data. This means that Bayesian estimation must be used rather than MLE. Both AMOS and Mplus use a form of ordinal regression, but LISREL uses linear
regression. Hence, a polychoric correlation matrix input for ordinal or dichotomous data must be used when using LISREL (Lee, 2007). The PRELIS program creates such a matrix for LISREL. Moreover, with numerical data or ordinal data with Likert categories greater than four, researchers often use correlations rather than covariances as input to estimation. Given this tendency, in most cases of maximum likelihood (as well as generalized least squares), fitting functions are scale invariant and the resulting estimates scale free. This has no effect on overall goodness-of-fit indices and parameter estimates (Bollen, 1989). If variables are standardized, however, standard errors may be inaccurate, meaning that regression paths and factor/error covariances can be erroneous or biased (Cudeck, 1989). Identifying whether or not a particular model is scale invariant is determined by a complex combination of elements, including how the factors are scaled (i.e., unit loading invariant [ULI] versus unit variance invariant [UVI]) and the presence of equality constraints on estimates of factor loadings (Kline, 2004). One case in which correlations should not be used is with multiple-group analyses. This is because when comparing models across samples, we are interested in evaluating absolute rather than relative effects (relative to differences in means and sample across groups); subsequently, unstandardized values should be used (Byrne, 2004). Browne (1982) proposed using the method of constraint estimation for correctly fitting a model to a correlation matrix instead of a covariance matrix. Constrained estimation involves imposing non-linear constraints on certain parameter estimates to guarantee that the model is scale invariant. These non-linear constraints, however, can be complicated to program manually (Steiger, 2001), and not all SEM programs support this type of constraint (e.g., LISREL, AMOS, EQS). The more specialized programs, such as SEPATH and RAMONA, allow constrained estimation to be programmed automatically.

All of the 21 papers examined used numerical and pseudo-numerical data with Likert categories of four and more. Moreover, covariances were specifically used in four studies (19%); correlation was employed in another four studies (19%); and 13 studies (62%) did not mention the analyses being based on covariances. This means it is possible that correlations were used as input (see Table 2). It is difficult to assess, however, how frequently the use of correlations has detrimental effects on the analysis. In future research, scholars should conduct all analyses on covariance matrices, and tests of significance for individual parameters should be reported from this analysis. Standardized parameter estimates, if needed, can be obtained easily from the standardized solution in which either the latent variables or both the latent and observed variables have been standardized.

**Sample size:** The effect of sample size is a more sensitive issue for some evaluative criteria of CFA than for EFA (Netemeyer et al., 2003). A small sample size causes non-convergence and improper solutions, such as negative variance estimates (Anderson & Gerbing, 1988; Boomsma, 1982). This makes parameter estimation impossible to interpret (Ding et al., 1995). Moreover, according to Netemeyer et al. (2003), although CFA sample sizes should be large, the “more is better” strategy might not always be appropriate. An excessive number of samples may show slightly significant differences between the observed and implied covariance matrices (or parameter estimates). The general recommendation is that the sample size be sufficiently large; that is, approximately 200 or more observations (Kline, 1998). Dillon, Kumar, and Mulani (1987) recommended a
sample size ranging from 100 to 150 to ensure the appropriate use of MLE. The number for a minimum ratio is at least five respondents for each estimated parameter (Joreskog & Sorbom, 1996). A ratio of 10 respondents per parameter, however, is considered most appropriate (Hair, Black, Babin, Anderson, & Tatham, 2010). Gorsuch (1983) suggested that the absolute minimum ratio is five individuals to each variable, but not less than 100 individual for any analysis. According to Anderson and Gerbing (1984), a sample size of 150 is usually enough to obtain a converged and proper solution for models with three or more indicators per factor.

The mean number of parameters estimated in the 21 papers examined was about 16 (see Table 2). The mean sample size was 148, resulting in a mean ratio of sample size to number of free parameters of about 9:1 for all papers. More specifically, the minimum ratio of parameters to sample size in the 21 papers reviewed was six, compared to the maximum ratio, which was 12 (see .2). These figures indicate that, on average, sample sizes considered in previous studies are broadly acceptable for obtaining trustworthy parameter estimates and valid tests of significance. Nevertheless, because it is easy to estimate beforehand the likely number of parameters, future papers should choose the necessary sample sizes that correspond to thresholds discussed herein. As Martin (1987) pointed out, there is often a trade-off between collecting high-quality data and gathering data from a large sample of respondents. A researcher’s primary objective, therefore, should be to obtain high-quality data even if SEM may not be an appropriate methodology to analyze that data (Bagozzi & Yi, 1989).

Model Identification

Another important consideration prior to conducting SEM analysis is to determine if the model to be estimated is identified. A model is identified if it is impossible for two distinct sets of parameter values to yield the same population variance-covariance matrix. A necessary condition for identifying a model is that the number of parameters to be estimated should be less than or equal to the number of distinct elements in the variance-covariance matrix of the observed variables. This implies that the number of degrees of freedom is non-negative. This rule is easy to check; unfortunately, however, it is not a sufficient condition for identification. General, easy-to-follow procedures to prove identification are available in specialized cases. For example, the two-indicator rule is a sufficient condition that applies to SEM models that feature unidimensional measurements and concern a minimum of indicators. An SEM model should have at least three indicators or two indicators with covariances among the factors to be identified. Models with only two indicators, however, are more prone to estimation problems (e.g., non-convergence and improper solutions). Bollen (1989) recommended assessing each latent variable with a minimum of three or four indicators. Furthermore, a significant number of observed indicators are more likely to tap all facets of the construct of interest. The drawback is that the greater the number of indicators per factor, the more difficult it will be to represent the measurement structure underlying a set of observed variables and find a model that fits the data well (Anderson & Gerbin, 1984).

Showing that a model is identified may be non-trivial for certain kinds of models. In particular, special care is required for models that are not unidimensional and/or non-
recursive (i.e., with feedback loops). Order and rank conditions represent, respectively, necessary and sufficient conditions to determine if a non-recursive model is identified. The order condition is met if the number of excluded variables for each endogenous variable exceeds the total number of endogenous variables minus 1 (Kline, 2004). The rank condition is usually described in matrix terms and is evaluated using the system matrix (Berry, 1984). Much, like the order condition, the rank condition must be evaluated for the equation of each endogenous variable. The rank condition can be viewed as a requirement, such that each variable in a feedback loop has a unique pattern of direct effects from variables outside the loop (Kline, 2004). Moreover, although a model can be theoretically identified, it still can remain unsolvable due to empirical problems such as high multicollinearity between factors and/or path estimates close to 0 in non-recursive models (Bollen, 1989). Among the 21 studies considered, none checked whether a model was identified prior to being estimated (see .3).

Identification was not considered explicitly because in many cases software programs warn the user if a model appears to be under-identified. Thus, identification problems might be detected even if identification is not proven theoretically (Baumgartner & Homburg, 1996). Moreover, identification can be avoided, a priori, by having more than three indicators per factor. Thus, the current study assessed the number of variables/indicators that reflected upon each factor. The mean of variables for all 21 studies examined was 3.5 (see Table 3). More specifically, the minimum mean of variables across the 21 studies was three, compared to the maximum, which was five. These results suggest no under-identification issues were evident.

Table 3: Statistics Concerning Identification Issues in Applying SEM to Tourism Demand Modeling

<table>
<thead>
<tr>
<th>Number of Papers (N = 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of the Model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Average number of manifest variables</td>
</tr>
<tr>
<td>Average number of latent variables</td>
</tr>
<tr>
<td>Average ratio manifest/latent</td>
</tr>
<tr>
<td>Min. average ratio manifest/latent</td>
</tr>
<tr>
<td>Max. average ratio manifest/latent</td>
</tr>
</tbody>
</table>

Overall Model Fit

When testing SEM models it is important to assess the overall fit of the observed data to an a priori model before assessing individual parameters (Joreskog et al., 1999). The most popular index for assessing the overall goodness of fit of a model has been the chi-square ($X^2$) statistic. Chi-square tests the null hypothesis that samples and fitted covariance matrices do not differ from one another (Hu & Bentler, 1999). One critical limitation of this statistic, however, is its known sample-size dependency (Kim & Yoon,
2004). If the sample size is more than 200, this measure has a greater tendency to indicate significance (Hair, Black, Babin, Anderson, & Tatham, 2010). The test is also suspect when using small samples because some are not distributed as chi-square populations (Brunham & Anderson, 1998). Interpreting chi-square, therefore, must be done with caution in most applications (Joreskog & Sorbom, 1996).

Another accepted index that uses test statistics to estimate how well the fitted model approximates the population covariance matrix per degree of freedom is the root mean squared error of approximation (RMSEA). Browne and Cudeck (1993) suggested that a value of RMSEA below 0.05 indicates close fit and that values up to 0.08 are reasonable. They also proposed a test of close fit to confirm the null hypothesis that RMSEA is equal to 0.05, against the alternative hypothesis that RMSEA is greater than 0.05. (In contrast, the conventional $X^2$ statistic tests the hypothesis that RMSEA = 0.)

To complement the chi-square and RMSEA statistical measures, the literature has encouraged using multiple criteria as ad hoc indices of fit (Byrne, 2001; MacCallum, 1986) because no “best” index exists (Yoo & Donthu, 2005). Some indices of fit are stand-alone indices that assess model fit in an absolute sense, including relative chi-square ($X^2$/df); the goodness-of-fit index (GFI); adjusted goodness-of-fit index (AGFI); and the root mean square residual (RMR). Others are incremental fit indices that compare the target model to the fit of a baseline model, including the Bentler and Bonett normed fit index (NFI); the Tucker and Lewis non-normed fit index (NNFI); and the normed comparative fit index (CFI). Rules of thumb, as opposed to significance tests, are used to determine acceptable fit levels because the underlying sampling distributions for these indices are unknown. For the relative chi-square, the recommended level is 2:1 (Kline, 1998). For GFI, AGFI, and CFI, the acceptable level is 0.90 (Hair et al., 2010). For the standardized RMR, the acceptable level is less than .010 (Schumacker & Lomax, 2004). For NFI and NNFI, the recommended level is .95 (Hu & Bentler, 1999). Thus, until a definitive measure is developed, researchers should use multiple measures to provide evidence about the overall fit of their models (Breckler, 1990).

The data in Table 4 shows that all published articles using SEM reported at least one stand-alone fit index (chi-square). Studies reported between two and nine fit measures, with an average number of five indices per paper examined. The most popular measure by far, as shown in .4, was chi-square (21 studies, 100%); followed by RMSEA (15 studies, 71%); GFI (13 studies, 62%); CFI (12 studies, 57%); and NFI (11 studies, 52%). These results indicate that researchers too often evaluate models on the chi-square test. Alternative fit indices—particularly those based on incremental fit as suggested by Gerbing and Anderson (1993)—should be used more often in future applications of SEM in order to better assess the overall fit of a proposed model to the researcher’s dataset.
Table 4: Statistics Concerning Overall Fit Issues in Applying SEM to Tourism Demand Modeling

<table>
<thead>
<tr>
<th>Overall Model Fit</th>
<th>Number of Papers (N = 21)</th>
<th>Frequency (N = 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square and significance</td>
<td>21</td>
<td>100%</td>
</tr>
<tr>
<td>Relative chi-square</td>
<td>8</td>
<td>38%</td>
</tr>
<tr>
<td>Goodness-of-fit index (GFI)</td>
<td>13</td>
<td>62%</td>
</tr>
<tr>
<td>Adjusted goodness-of-fit index (AGFI)</td>
<td>10</td>
<td>48%</td>
</tr>
<tr>
<td>Root mean square residuals (SRMS residuals)</td>
<td>6</td>
<td>29%</td>
</tr>
<tr>
<td>Comparative fit index (CFI)</td>
<td>12</td>
<td>57%</td>
</tr>
<tr>
<td>Normed fit index (NFI)</td>
<td>11</td>
<td>52%</td>
</tr>
<tr>
<td>TLI, also called the Bentler-Bonett non-normed fit index (NNFI)</td>
<td>10</td>
<td>48%</td>
</tr>
<tr>
<td>Root mean square error of approximation (RMSEA)</td>
<td>15</td>
<td>71%</td>
</tr>
<tr>
<td>Average number of indices used in each paper</td>
<td>5</td>
<td>24%</td>
</tr>
<tr>
<td>Minimum number of indices used in a paper</td>
<td>2</td>
<td>10%</td>
</tr>
<tr>
<td>Maximum number of indices used in a paper</td>
<td>9</td>
<td>43%</td>
</tr>
</tbody>
</table>

Reliability and Validity

Measures of reliability and validity must be assessed when using SEM. A reliability value above 0.7 is regarded as satisfactory (Nunnally & Bernstein, 1994). The most common measure of reliability is the Cronbach’s alpha. This measure, however, has several limitations. For example, the coefficient alpha wrongly assumes that all items contribute equally to reliability (Bollen, 1989). Thus, a better choice is composite reliability, which measures reliability based on standardized loadings and measurement error for each item (Bollen, 1989). The Average Variance Extracted (AVE) is another measure of reliability that reflects the amount of variance in the indicators accounted for by a construct. Higher AVE occurs when indicators are truly representative of the latent construct. The AVE value should exceed .50 for a valid construct (Fornell & Larcker, 1981).

In coding whether an examined study described reliability and what measures were reported, it was determined that reliability was examined in 12 studies (57%). Nine studies (43%) disregarded reliability. Among the 12 studies in which reliability was examined, all used Cronbach’s alpha (see Table 5). In six studies (29%), Cronbach’s alpha was used with another measure such as composite reliability or AVE. Overall, AVE was reported in six studies (29%), while composite reliability was reported in three studies (14%).
Table 5: Statistics Concerning Reliability and Validity Issues in Applying SEM to Tourism Demand Modeling

<table>
<thead>
<tr>
<th>Constructs Reliability</th>
<th>Number of Papers (N = 21)</th>
<th>Frequency (N = 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes/No</td>
<td>Yes=12 No=9</td>
<td>Yes=57% No=43%</td>
</tr>
<tr>
<td>AVE</td>
<td>6</td>
<td>29%</td>
</tr>
<tr>
<td>Cronbach’s alpha</td>
<td>12</td>
<td>57%</td>
</tr>
<tr>
<td>Composite reliability (Goldstein Rho)</td>
<td>3</td>
<td>14%</td>
</tr>
<tr>
<td>At least one</td>
<td>12</td>
<td>57%</td>
</tr>
<tr>
<td>At least two (Cronbach’s alpha with AVE or Goldstein Rho)</td>
<td>6</td>
<td>29%</td>
</tr>
</tbody>
</table>

Convergent Validity

<table>
<thead>
<tr>
<th>Yes/No</th>
<th>Yes=11 No=10</th>
<th>Yes=52% No=48%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loadings + Significance</td>
<td>10</td>
<td>48%</td>
</tr>
<tr>
<td>AVE</td>
<td>7</td>
<td>33%</td>
</tr>
<tr>
<td>Cross-Loadings</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>At least one</td>
<td>11</td>
<td>52%</td>
</tr>
</tbody>
</table>

Discriminant Validity

<table>
<thead>
<tr>
<th>Yes/No</th>
<th>Yes=7 No=14</th>
<th>Yes=33% No=67%</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVE versus Communalities</td>
<td>4</td>
<td>19%</td>
</tr>
<tr>
<td>Confidence interval (Correlation + or - 2 std error)</td>
<td>2</td>
<td>10%</td>
</tr>
<tr>
<td>Chi-square difference between competing models</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>At least one</td>
<td>7</td>
<td>33%</td>
</tr>
</tbody>
</table>

Two validity subtypes should be further examined: convergent validity and discriminant validity. AVE is a suggested criterion of convergent validity (Fornell & Larcker, 1981). An AVE value of at least 0.5 (i.e., a latent variable is able to explain, on average, more than half of the variance of its indicators) indicates sufficient convergent validity (Gotz, Liehr-Gobbers, & Krafft, 2010). Examining the significance of the individual item loadings and inspecting the factor loadings for the observed variables is also suggested; absolute standardized loadings should be higher than 0.7 (Churchill, 1979). In the 21 studies examined, convergent validity was examined in 11 studies (52%), with 10 studies (48%) disregarding convergent validity. Among the 11 studies in which convergent validity was examined, seven studies (33%) reported AVE. Less stringent assessments of
convergent validity (e.g., examining factor loadings and significance) were used in 10 studies (48%).

Calculating the shared variance between two constructs and verifying that the result is lower than the AVE for each individual construct is the most common way to assess discriminant validity in the measurement model (Fomell & Larcker, 1981). Another way to examine discriminant validity is to check that the considered confidence intervals of the correlations between constructs do not contain 1 (Bollen, 1989). Yet another viable approach is to apply the chi-square difference test (Bagozzi et al., 1991) between a competing convergent model (one in which the correlation between the constructs is set at 1) and a discriminant model (one in which the observed correlation between the constructs is freely estimated). A significant difference would suggest that the two constructs do not overlap; that is, they are divergent.

In the examined studies, discriminant validity was examined in seven studies (33%), while 14 studies (67%) disregarded discriminant validity. Among the studies in which discriminant validity was examined, four studies (19%) reported AVE compared to shared variance; two studies (10%) cited the examined the correlation matrix; and one study (4%) reported pairwise tests. Table 5 reports this data. These results for reliability and validity are cause for concern because when measures are unsuspectingly weak, the tests are less powerful. The consequences may include mis-specifying the constructs and subsequently inappropriately modifying structural models and thus spurious findings. Low reliability and validity may cause relationships to appear non-significant, regardless of whether the links exist. This error would cause pertinent research paths to be prematurely cut short. Thus, researchers should carefully attend to the issues of reliability and validity of measurement constructs in future studies.

**Model Respecification**

If the fit of the implied theoretical model is not as strong as preferred, the model must be respecified. The respecified model is then subsequently reevaluated to verify if it fits the data better (Schumacker & Lomax, 2004). Various statistical tools are available to trace model misspecifications including modification indices and residual analysis. SEM output routinely reports this information. Respecification, however, must always be guided by substantive meaning. For example, Anderson and Gerbing (1988) argued that respecifications should be based on theory and content considerations to avoid exploiting sampling error, thereby achieving satisfactory goodness-of-fit. Chin (1998) and Kelloway (1995) asserted, however, that it is not enough for respecification to be theoretically justified; instead, respecified models should be validated with a new sample. Indeed, Brannick (1995) argued that respecifications should not be undertaken at all. If a theoretical justification for a modification exists, alternative models should have been proposed *a priori* rather than making *a posteriori* changes.

Among the 21 studies examined, 18 (86%) included a respecification (see Table 6). One study (5%) explicitly illustrated the respecifications, but none of the studies validated changes in the specification on a holdout sample. In 13 studies (64%), the authors cited theoretical reasons to justify the changes. Knowing that the majority of models reported in the tourism demand literature have been modified at least to some extent, it is critical for
future studies to be explicit when reporting how the researchers arrived at their final model. This will enable readers to understand to what degree the initially proposed model was altered and whether cross-validation was needed. Further, each respecification proposed should have substantive grounding to ensure that the literature supports the modified model. Finally, validation of the respecified model on holdout samples should be adopted whenever possible (i.e., whenever the sample size is large enough). When the sample is not large enough, other alternatives can still be used (such as bootstrap resampling techniques) to generate several samples on which the model can be validated. Future studies should use these suggestions to ensure the findings obtained through a respecified model are replicable, because failing to do so may mean the results remain doubtful.

Table 6: Statistics Concerning Model Respecification and Reporting in Applying SEM to Tourism Demand Modeling

<table>
<thead>
<tr>
<th>Model Respecification</th>
<th>Number of Papers (N = 21)</th>
<th>Frequency (N = 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes/No</td>
<td>Yes = 18</td>
<td>86%</td>
</tr>
<tr>
<td></td>
<td>No = 3</td>
<td>14%</td>
</tr>
</tbody>
</table>

Reporting

Kline (2004) contended that articles should provide enough detail to permit others to duplicate the results, because replication serves as a precaution against accepting erroneous findings. Chin (1998) suggested that the input matrix, the computer program and version used, the models tested, computational options used (starting values, number of iterations, estimation methods, etc.), and anomalies encountered during the analytical process should be explained clearly in any SEM paper. Such details are vital. For example, different programs and versions of the same program can generate different default parameters, thereby misleading the reader (Hoyle, 1995). In addition, starting values may lead to non-convergence of the results or failure of iterative estimation (Kline, 2004). To correct this, the researcher needs to select start values that differ from the previous run. Bentler (1995) provided guidelines for calculating the correct start values. Researchers should communicate the steps followed so that third-party reviewers can replicate the results.

Among the 21 studies evaluated, no single study listed all of the items specified (see Table 6). Six studies (29%) reproduced the input matrix, but the remaining 15 (71%) did
not. The software package used was reported in only nine studies (43%). LISREL was used in five studies (24%); AMOS in two studies (10%); and CALIS and Mplus were used in one study each (5%). In all cases, the version of the software used was reported if the software package was named (nine studies, 43%). Finally, anomalies encountered were only reported in one study (4%), where the researchers encountered non-convergence due to high multicollinearity among the variables. These results suggest that many of the reported studies lack significant usefulness for future research and replicating results. Thus, future work on tourism demand that uses SEM must share these details so that readers gain a more comprehensive understanding of the analysis performed. Doing so will allow researchers to replicate results in succeeding works.

Conclusions

Applying SEM is still limited in tourism demand modeling, although at least two features make SEM an attractive candidate for analyzing tourism data. First, SEM allows the researcher to assess latent constructs explicitly and correct for unreliable measures, provided multiple indicators of each construct are available. Second, SEM makes it possible to investigate, using a simple approach, comprehensive theoretical frameworks in which the effects of constructs are propagated across multiple layers of variables via direct and indirect paths of influence. These advantages, coupled with developing more sophisticated, yet user-friendly computer programs to estimate and test such models, make SEM a solid approach for widespread use in studying tourism demand.

Based on reviewing applications of SEM in seven journals over a 12-year period, this study explained how SEM was applicable to tourism demand research from a technical perspective and examined how well (or not so well) these applications have been implemented. In particular, it was found that the vast majority of the papers (18, 86%) failed to specify how the issue of missing data was handled; the remaining three papers used widely different methods to address the problem. In addition, results indicated that few papers (2, 9%) discussed the distribution of the data (normality), while all used traditional estimation methods to fit the model, which routinely assumes normality of the dataset; subsequently, a priori support of normality is lacking. Moreover, in assessing the fit of the model, findings indicate that the totality of the papers could use more alternative fit indices to further support test statistics and model fit, with the test statistics being sensitive to sample size and complexity of the constructs. Results also showed that only half of the studies (12, 51%) tested for the reliability and validity of the individual constructs, knowing that not doing so this could lead to error in specifying the constructs and making the tests less powerful. Lastly, results showed that the vast majority of the papers (18, 86%) listed a respecification of the model without providing a theoretical justification for this modification. Furthermore, very few studies provided enough details (e.g. input matrix - 6, 29%; software uses - 5, 24%; etc.), to permit replication of the studies by future readers. Finally, and in contrast to the previously highlighted shortcomings, results indicated that the sample size used in the totality of the studies was large enough as to obtain trustworthy and valid results from the model.

In spite of this last finding, results indicate that although SEM is being used, it tends to be used inappropriately, and thus that there is room for improvement in the application of
SEM and its presentation in papers on forecasting tourism demand To assist in this process, we have created a checklist of methodological guidelines for use in tourism demand forecasting (see Table 7). These guidelines are based on prior discussions and the present review of published studies. Reviewers and researchers can use this checklist with future work to maximize SEM’s utility. Guidelines are as follows. First, when data are collected, screen carefully for missing values, normality, and linearity before a covariance matrix is computed and models are investigated. Second, give careful thought to model identification before empirical data are collected; identify both CFA and SR models if using a two-step approach. Experience indicates that SEM is most constructive for theoretical frameworks of moderate complexity in which each construct is measured by an average set of indicators. Third, regarding model estimation and testing, assess the performance of a given model (either CFA or SR) specification using a variety of global and incremental fit measures. Fourth, check individual measures of reliability and validity. This is traditionally done at the CFA model level (if a two-step approach is used). Fifth, consider alternative theoretical models when possible for both CFA and SR, and cross-validate research results when possible. Finally, because the discipline desires accumulated knowledge, provide adequate information about statistical procedures to guide future researchers.

As with any study, this study has limitations. One major limitation is that it assessed applying SEM based on information reported in published articles. In some cases, authors may have made appropriate decisions and discussed decisions with referees during the review process, but did not include such material in the article. Another limitation is the small number of articles examined. As the use of SEM becomes more common in tourism forecasting, a more comprehensive review can be prepared.

Until then, this first review of prior applications of SEM, however, is anticipated to improve the quality of empirical research in tourism forecasting. Significant opportunities remain for using SEM to generate insight into tourism demand modeling at both the aggregate and individual levels, and this study provides a set of guidelines to follow when this methodology is used in the domain. This will certainly prove helpful for future studies in the discipline and could motivate future researchers to use SEM without becoming concerned regarding its proper use.
Table 7: Checklist of Issues to Consider in SEM Studies

1. Sample issues
   (a) Sample size.
   (b) Number of observations per parameters.
   (b) Missing data (evaluate pattern and how to correct for missing values).
   (c) Distribution of sample and linearity (check for normality and linear correlation).
   (d) Type of data (numerical/categorical, correlation/covariance matrix).

2. Identification
   (a) Degrees of freedom.
   (b) Number of indicators per construct (for recursive models).
   (c) Order and Rank Matrix conditions (for non-recursive models).

3. Overall Model Fit
   (a) Statistical tests (exact fit versus close fit).
   (a) Mix of absolute and relative fit indices.

4. Measurement issues
   (a) Reliability of measures.
   (b) Measures of discriminant validity.
   (c) Measures of convergent validity.

5. Modification
   (a) Respecified models given status of hypothesized (modifications supported by literature).
   (b) Validating changes in cross-validated model.

6. Reproducibility issues
   (a) Input matrix.
   (b) Name and version of software package used.
   (c) Computational options used (starting values, estimation method, etc.).
   (d) Analytical anomalies encountered.
References


Guy Assaker
Tel.: +961 1 786456
E-mail: gassaker@lau.edu.lb