Tourism Management 68 (2018) 465–469

Contents lists available at ScienceDirect

Tourism Management

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

Reflective versus formative confusion in SEM based tourism research: A critical comment

Josip Mikulić[a,*], Chris Ryan[b]

[a] Faculty of Economics and Business, University of Zagreb, JF Kennedy Square 6, 10000, Zagreb, Croatia
[b] China-New Zealand Tourism Research Unit, University of Waikato Management School, Hamilton, New Zealand

ARTICLE INFO

Keywords:
Formative
Reflective
Measurement mode
Misspecifications
Structural equations modeling

ABSTRACT

The aim of this note is to raise awareness about the problem of measurement misspecification in tourism research with reference to studies that employ structural equations modeling. More precisely, this note describes possible instances when the measurement mode has been wrongly specified as reflective, although employed indicators/measures are formative by their nature. To assist future tourism research based on latent variable modeling, this commentary provides pragmatic advice and points to some of the key works from the wider marketing and management literature.

1. Introduction

Structural equations modeling (SEM) is a popular technique in tourism research. What has made SEM so popular is its ability: (i) to simultaneously test numerous relationships between theoretical constructs, which are measured as latent variables as identified by several manifest indicators and (ii) to regard some of these variables as having mediating or moderating roles when seeking to explain specific behaviors or perceptions of actions.

The number of SEM-based applications in tourism research began to accelerate in the years after the introduction of AMOS and subsequently SmartPLS (Ringle, Wende, & Will, 2005). These packages owed much of their popularity to their user-friendliness, as they were significantly more intuitive and easier to handle than, for example, LISREL, which at that time required the entry of code. Researchers could simply draw the diagrams representing the theoretical relationships being examined – a feature that has now been copied by several other pieces of software including EQS, LISREL, Stata and MPlus.

The introduction of SmartPLS was also important, as unlike the covariance based regimes of the other packages, the use of partial least squares represented an alternative approach to modeling relationships. Covariance based structural equation modeling (CB-SEM) considers "the constructs as common factors that explain the covariation between its associated indicators" (Hair, Hult, Ringle & Sarstedt (2017, p. 15), whereas partial least squares structural equation modeling (PLS-SEM) creates (a) weighted composites of indicator variables and (b) proxies for constructs with the assumption of a common factor being relaxed. This meant that the assumptions that underlie any form of regression, namely that normality of distribution exists, that there is independence between variables and there is a properly constructed random sample (Hair et al., 2017), need not be rigidly adhered to. In tourism research where these assumptions are more often absent than present, the PLS-SEM approach offered a means of escaping these traditional rigidities associated with CB-SEM. For example, given that most holiday-makers would select a holiday expected to provide more satisfaction than dissatisfaction, data seeking to record evaluations of place or experience would often be heavily negatively skewed if not showing high levels of kurtosis. PLS-SEM also offered other advantages in that conventionally CB-SEM was treated as a method requiring large samples, whereas PLS-SEM could be undertaken with smaller sample sizes (Hair et al., 2017). However, a survey of the literature indicates some variance as to what constituted a 'large sample'. Hair, Anderson, Tatham, and Black (1998, p.605) were often cited as indicating that 200 respondents represented a large sample, whereas Westland (2010a,b) devised an algorithm that generally indicated that such a size was far too small given the number of latent and observed variables being calculated, while in addition many researchers additionally failed to take into account the power and size effects and their implications for sample size (Ellis, 2010, pxiv-xv).

In the view of the authors, in practice many researchers would simply ignore these issues, and indeed would ignore other general conventions, coming to treat confirmatory factor analysis as in practice an exploratory technique. Indeed, recently in various presentations Ringle has suggested that these conventional reasons for the use of partial least squares has over-shadowed what he perceives as a significant reason for the use of Smart-PLS; namely that it has value as an exploratory form of analysis as distinct from being confirmatory. As

* Corresponding author.
E-mail address: jmikulic@efzg.hr (J. Mikulić).

https://doi.org/10.1016/j.tourman.2018.05.002
Received 3 March 2017; Received in revised form 22 March 2018; Accepted 9 May 2018

Awang, Afthanorhan, and Asri (2015, p. 59) state:

However, not many … know that the analysis in VB-SEM (variance-based SEM) is only meant for exploratory study as opposed to confirmatory analysis in CB-SEM (Hair, Sarstedt, Hopkins, & Kuppelwieser, 2014). There are great differences between the types of analysis from the statisticians' point of view. According to (Hair et al., 2014), the algorithm employed in VB-SEM or popularly known as PLS-SEM (Smart-PLS and Warp-PLS) is Generalized Least Squares (GLS) while the algorithm employed in CB-SEM (Amos, etc.) is the Maximum Likelihood Estimator (MLE). These two types of algorithm differ greatly in term of efficiency of their statistical estimates for path coefficients.

Hence it is the view of the authors that a danger existed where researchers used PLS-SEM as an exploratory technique but labored under the misapprehension that it still needed to be used as a confirmatory technique. Moreover, as Nunkoo, Ramkissoon, and Gursoy (2013) noted, when conducting CB-SEM studies, tourism researchers frequently failed to adopt best practice guidelines available from the wider business and psychology literature. Additionally, it is not uncommon to find that a requirement for separate samples for exploratory and confirmatory factor analysis was ignored, and in the search for higher order modeling the impression was formed that researchers sought to fit the data to meet a proposed hypothesis rather than being prepared to report data that failed to support pre-conceived models. In noting this, it is recognized that the need for separate samples remains contested and authors such as Marsh et al. (2009) argue strongly for integrated approaches.

Another common failing was that while indices of fit were being reported, there was often a silence over coefficients of determination, or alternatively low coefficients were reported without comment. It is recognized that "causal paths in a structural model are negligible" (Reichardt, 2002, p. 310), but if the PLS-SEM is essentially exploratory in nature, to disregard the statistic seems contrary to the exploratory process. It comes back to the issue of the research design and the purpose of the research, a point again explored below.

This silence also extended to other omissions, and a significant one in the eyes of the authors was the all too common silence about the descriptive statistics. These were often not reported, yet as noted above about the nature of score distributions, they possess important information. In one case known to the second author a researcher presented a SEM calculation. However, when being pushed to provide the actual mean scores it then became obvious that the major finding had been totally ignored. The majority of the scores were at the mid-point of the scale, thereby implying that the majority of the sample was simply indifferent to the experiences being evaluated. A finding which was totally overlooked in the rush "to press the buttons for AMOS", thereby illustrating what the second author has termed the phenomenon where the researcher "turns the computer on and turns the brain off". The descriptive statistics clearly indicate what respondents believe is or is not important as discussed below.

Other serious issues also exist. In particular, when examining many of the SEM-based studies in the tourism literature, whether using a variance- or covariance-based approach, there seems to persist a rather widespread misunderstanding or confusion with regard to the principles of underlying measurement theory. More precisely, the choice of whether a formative or a reflective measurement operationalization should be pursued when modeling theoretical constructs seems to be causing significant trouble. It is the purpose of this paper to state that, especially if an exploratory approach is being adopted, it is essential to understand the distinction between formative and reflective measures.

2. Reflective versus formative destination image

To illustrate much of the problems in tourism-related SEM research based on latent variable modeling, one can consider 'destination image' as an example of a widely studied theoretical construct in the tourism literature. When put into quotation marks and combined with 'SEM' or 'Structural Equation(s) Model(l)ing', SCOPUS returns 75 journal articles that bear these terms in their title, abstract or keywords (accessed on March 1st, 2017). If taking a less stringent, but also less precise search approach on Google Scholar (not restricted to title, abstract, and keywords), then several thousand contributions are returned.

An examination of the 75 articles returned by SCOPUS reveals that 66 (88%) of them use a reflective measurement operationalization for destination image, while three (0.4%) use formative operationalization. Two more studies use single-item operationalization (i.e. overall image assessment), while four studies do not model the destination image construct at all. Out of the 66 reflective approaches, 16 use a hierarchical operationalization after having conducted an exploratory factor analysis. Hereafter, the construct is then modeled reflectively at both first and second order.

The fact that not all these studies use the same approach for measuring destination image (i.e., either reflective or formative) should not be regarded as problematic *per se*, because it is the nature of indicators that determine the proper (i.e. applicable) mode of measurement. If indicators reflect or are a consequence of the image of a destination (i.e. causality flows from destination image to the indicators), then only a reflective measurement mode is applicable. Conversely, only a formative mode is applicable if indicators form or cause the image of a destination (i.e. causality flows from the indicators to destination image (see e.g. Edwards & Bagozzi, 2000; Edwards, 2010; Jarvis, MacKenzie, & Podsakoff, 2003; Podsakoff, Chen, & Podsakoff, 2006).

Another way of posing the issue is provided by Coltman, Devinney, Midgley, and Veniak (2008) who simply specify that in any model the formative model is one constructed by the researcher, whereas in the reflective the latent construct has a prior existence. Alternative terminology for such directions of causality are that the reflective is an effect model and the formative is a causal model. In this case it has an existence independent of the measures used and hence one looks for the consequences of its existence. In the formative model the latent variable is calculated from the observable measures. In terms of directions of causality, in the formative model the direction of causality moves from the items to the construct, and in the reflective the opposite is true. The statistical consequence of this is that in the reflective model the correlations between the measures should be high and conventional measures such as the Cronbach alpha coefficient should indicate high degrees of reliability. On the other hand the formative model constructed by the research may lack such high scores, yet nonetheless should be in a predicted direction. Because of this, some researchers will use a Multiple Indicators and Multiple Causes (MIMIC) Model or seek other means of measuring structural linkages.

Although these rules are rather straightforward, specification of the proper measurement mode seems to be causing confusion in many studies taken from the tourism literature. At the same time, however, proper measurement specification is a crucial requirement for valid creation of empirically grounded tourism theory. For example, consider the indicators *Accessibility of the destination*, *Variety and quality of accommodations* and *Cultural and historical attractions*. These three indicators, among several others, have been used in a reflective operationalization of destination image in an article published in a renowned tourism journal. Many other articles from other journals take a similar approach. If we assume that these three indicators were truly reflective indicators, then the bottom line assumption would be that all the indicators are strongly correlated, because they reflect the same construct—i.e. destination image. Theoretically, a single, perfect reflective indicator would, in fact, do, but finding two, three or more indicators which are highly correlated, gives us confidence (i.e. reliability) that what is being truly measured is indeed what the researchers intended to measure—i.e. the so-called 'useful redundancy' is achieved (DeVellis, 1991).

Accordingly, if the three indicators truly were reflecting the construct "destination image", then these indicators would (i) highly load on the underlying construct and (ii) a high Cronbach alpha coefficient

would be obtained, so signaling internal consistency and construct reliability. Moreover, many studies take this latter result as a sufficient signal that also (iii) construct validity has been achieved. Put differently, strong correlation among these three indicators is regarded as proof that we really measured destination image. In short, in itself the SEM software cannot draw the causal patterns and directions of arrows; but rather theoretical insights and evaluations remain to be made by the researcher.

3. An illustrative example: Machu Picchu versus Paris

Without excessive elaboration, it is, however, rather clear that these three indicators do not necessarily have to be mutually correlated for every destination. Consider, for example, Machu Picchu. The destination is not easily accessible, it does not have a variety and quality of accommodations, though it does certainly have breathtaking cultural and historical attractions. In contrast, Paris would likely achieve consistently good ratings on each of these three indicators.

Now, does this make a difference? Yes, it makes a big difference! The difference is that the reflective, three-indicator measurement operationalization for the theoretical construct destination image would obviously not work well if applied to measure the destination image of Machu Picchu. At least one indicator would not be strongly correlated with any of the others. This, in turn, would lead to a low loading of the uncorrelated indicator, which is rather mechanically interpreted in many studies as meaning that the indicator does not reflect the underlying construct and thus needs to be discarded (although reliability and validity of the whole construct operationalization has been found satisfactory in an earlier study). When discarded, only the correlated indicators remain, internal consistency rises and everything appears to be empirically fine. In the case of Paris, however, there would likely be no need to drop any of the indicators. Consequently, we would find ourselves using different measurement operationalizations to measure the same theoretical construct—i.e. destination image. Rather more precisely, we would be using only that part of the measurement operationalization that fits the data well to measure the image of Machu Picchu, while we would be using the whole operationalization to measure the image of Paris.

So, does this makes sense? Of course, it does not if seeking to establish a general model of destination image! In the above illustrative example of destination image, what one arrives at when dropping indicators is, in fact, only spurious construct validity, signaled by forced internal consistency and construct reliability that has only been achieved by discarding indicators that did not behave well within the data. Many studies from the tourism literature proceed this way, thereby departing from construct operationalization found reliable and valid in previous studies. To be truly valid, however, any reflective measurement operationalization needs to be robust across different cases, since the elementary assumption of reflective measurement is that indicators are mutually correlated and thus exchangeable one with another (Jarvis et al., 2003).

What went wrong? What went wrong in this illustrative example is that the measurement mode has been mis-specified as reflective, even though it could have only been specified as formative given the nature of the used measures. To be clear, the image of a destination does not cause good or bad accessibility, more or less diverse accommodation options, or high or low cultural value. Rather obviously, it is the other way round. This becomes clear once the perspective is shifted away from the realm of empirical reliability and validity assessments, to how things are logically related to each other in the real world. However, because of the widespread misbelief that a reflective operationalization should generally apply or be preferred when using SEM, many studies blindly obey the reliability imperative in a classical psychometric manner (i.e. the Cronbach alpha principle), without having made an *a priori* examination whether a reflective measurement mode is actually applicable given the nature of measures used. To make it worse, because the measurement mode has been mis-specified as reflective instead of formative, many studies rather nonchalantly discard indicators which are, in fact, really relevant for measuring destination image.

Another issue that needs to be borne in mind is the suspicion that many researchers are tempted by a perceived need to report good indices of fit. Bentler (2007), in a debate over the use of indices suggested that researchers should be required to "submit a separate statement that verifies, for each major model, that (a) every parameter in the model is purely *a priori*, and if not, (b) details on all model modifications that were made. This material should be sent to reviewers along with the manuscript" (p.825) – something that in the experience of the authors very rarely happens. Indeed, in the experience of the second author attempting to obtain papers that comply with the Principles of the Committee of Publication Ethics (COPE) is difficult to attain due to a tolerance of less than perfect reporting previously. As an aside it can be said that too many papers still adhere to a practice of reporting variables with the use of acronyms that fail to disclose the actual items used in a questionnaire.

It is also suggested that in thinking about formative and reflective models researchers should consider the practice suggested by Baron and Kenny (1986) whereby they follow a four-step practice of regression testing for testing mediating variables, which process in itself breaks down or disaggregates the patterns of relationships, thereby, in the view of these authors, encouraging thinking about the relationships being tested, and whether they are indeed reflective.

Often this temptation to discard indicators also emerges from a failure to actually consider the raw data, and more specifically a failure to look at the descriptive statistics. It has been previously noted that often in tourism research psychological data pertaining to attitudes, evaluations and perceptions tend to be negatively skewed, and thus the use of PLS-SEM can overcome that issue. Researchers however can also ignore other issues, one being the failure to consider missing data. Two problems arise here. The first is simply a failure to provide an option for a non-response. This has the advantage of retaining all the respondents' scores, but now includes the possibility of mis-representing the reality of the data set. Hence, if for example, a questionnaire relating to meals in a restaurant includes an item such as an assessment of the waiter's ability to recommend a wine, how does the teetotaler respondent reply? If the option has been included, how then are the missing data handled? Consequently this journal requires copies of the questionnaire to be submitted with the manuscript.

Examination of the descriptive statistics may also provide clarification as to whether the used measurements actually fit a reflective or formative operationalization, with prior steps such as the use of correlation or regression analysis helping to formulate an appropriate response. Various other issues can also come to light, one being the probability of spurious correlations that may well provide artificially high weightings in any SEM approach. As briefly noted above, in one instance a researcher used a seven point scale, and on being asked for the descriptive statistics it proved that almost all of the scores varied around a value of 4.0 with relatively little standard deviation. On being informed that the major finding was that the respondents were generally indifferent as to the suggested items, the editor obtained the response that 'such a thing had not been considered'. High scores had been achieved in the indices of fit, partly because of high spurious correlations across patterns of scores due to a relative absence of variance.

4. Conclusion

The motivation for writing this note was to point to the perceived problem of measurement misspecification in the tourism research area. More specifically, this note is most concerned with the increasing number of SEM-based studies that force formative indicators into a reflective construct operationalization, accompanied by a failure to

actually consider the patterns of scores and the content of the questionnaire items being used. When modeling theoretical constructs, tourism researchers are thus strongly advised to critically examine the applicability of a reflective or a formative measurement mode based on the nature of measures used. It is suggested that a simple first stage is to actually examine elementary descriptive statistics such as mean scores, measures of dispersion and correlations. Such a simple step will indicate what respondents perceive as important or highly ranked, and equally what is not. Measures of dispersion indicate degrees of variance within the sample while, of course, correlations will indicate potential latent relationships. Taken together they may already indicate whether a reflective or formative relationship between measures and targeted constructs exists.

In order to minimize potential confusion in the first place, it is, however, strongly recommended that researchers should specify the measurement mode for targeted constructs before designing questionnaires and defining measures, and then adhere to guidelines available in works from the general marketing, management and psychology literature (e.g. Bollen, 2007; Edwards & Bagozzi, 2000; Jarvis et al., 2003). In short, issues of research design need to be carefully considered.

Another implication emerging from this brief discussion is the ever-growing need for consensual agreement upon both construct definitions and subsequent measurement operationalizations in order to create valid and robust tourism theory. It is, in fact, impossible to replicate studies and verify theories in a reliable and valid manner if using different measurement operationalizations from case to case. To state the obvious, case based construct reliability and validity has no value beyond the case being analyzed, and indeed may not possess even that if the construct operationalization is not replicated in other studies. It is for this reason the journal asks for questionnaires to be appended, so permitting the possibility of replication in other contexts. However, even such replication may be misleading if all that is replicated is a misconstructed model – in short it is rather easy to justify one's own wrong measurement approach by referring to someone other's wrong approach published in a good or even top tourism journal as measured by impact factors. The first thing a researcher should be looking for is whether an apparently reliable and valid measurement from one study needs to be revised in a second study to retain such reliability and valid – this is a warning sign that all is not well.

To conclude with an easily comprehensible example, an unreasonably stripped-down two-item reflective measurement operationalization, caused by an initial misspecification of measurement mode, could enable researchers to empirically prove that the image of a Yugo GV is *better* than the one of a Lamborghini Diablo, only because the Yugo has rear seats and a trailer hitch, while the Lamborghini has not. Once this has been empirically proven, published and replicated, it becomes, however, an increasingly difficult mission to prove that this is simply wrong, and that this result is due to an elementary misunderstanding of psychometrics and measurement theory. Obviously, the progress of theory building may not only be hindered, but it may also be regressing given a misspecified model like this one. Likewise, practical implications from such models may be severely misleading.

At the 2017 Conference of the International Academy for the Study of Tourism Smeral (2017) argued that as researchers, tourism academics have become seduced by the necessity of fitting models and results into what may be described as a strait jacket of statistical parameters and measures of significance. He suggested that researchers have not only to test, but also to estimate; in short to look more carefully at the social parameters within which individuals and communities work. Much of the same perspective can be applied to those publications that utilize structural equation models.

Acknowledgement

This work has been supported by the Croatian Science Foundation under the project DESTBRAND (project no. UIP-2014-09-7005, Period: 2015–2018).

References

Awang, Z., Afthanorhan, A., & Asri, M. A. M. (2015). Parametric and non parametric approach in structural equation modeling (SEM): The application of bootstrapping. *Modern Applied Science, 9*(9), 58–67.
Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological Research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.
Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42*(5), 825–829.
Bollen, K. A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods, 12*(2), 219–228.
Coltman, T., Devinney, T. M., Midgley, D. F., & Veniak, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research, 61*(12), 1250–1262.
DeVellis, R. F. (1991). *Scale development: Theories and applications.* Newbury Park, CA: Sage.
Edwards, J. R. (2010). The fallacy of formative measurement. *Organizational Research Methods, 14*(2), 370–388.
Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*(2), 155–174.
Ellis, P. D. (2010). *The essential guide to effect Sizes: Statistical power, meta-analysis, and interpretation of research results.* Cambridge: Cambridge University Press.
Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis.* Upper Saddle River, NJ: Prentice Hall.
Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., & Sarstedt, C. (2017). *A primer on partial least squares structural equation modelling (PLS-SEM).* Thousand Oaks, Calif: Sage.
Hair, J. F., Jr, Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): an emerging tool in business research. *European Business Review, 26*(2), 106–121.
Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*(2), 199–218.
Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., et al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of University Teaching. *Structural Equation Modeling A Multidisciplinary Journal, 16*(3), 439–476.
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.
Podsakoff, N. P., Shen, W., & Podsakoff, P. M. (2006). The role of formative measurement models in strategic management research: Review, critique, and implications for future research. *Research methodology in strategy and management* (pp. 197–252). Emerald Group Publishing Limited.
Reichardt, C. S. (2002). The priority of just-identified, recursive models. *Psychological Methods, 7*(3), 307–315.
Ringle, C. M., Wende, S., & Will, A. (2005). *SmartPLS 2.0 (beta)*.
Smeral, E. (2017). Substantive versus statistical Significance: Contradiction or coexistence in tourism. *International Academy for the study of tourism – 15th biennial conference. "Tourism in Turmoil"*Guangzhou, China: Sun Yat Sen University 25th May.
Westland, J. C. (2010b). Erratum to "Lower bounds on sample size in structural equation modeling" [Electron. Commerce Res. Appl. 9 (6) (2010) 476–487]. *Electronic Commerce Research and Applications, 11*(4), 445.
Westland, J. C. (2010a). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications, 9*(6), 476–487.

Josip Mikulić is Associate professor and Head of the department of tourism at the Faculty of Economics and Business, University of Zagreb, Croatia. His research focus is on tourism marketing management, destination branding and research methods in business and economics. He has published in most of the leading tourism journals and received several awards for his research.

Chris Ryan is Professor of Tourism at the University of Waikato Management School, and director of its China-New Zealand Tourism Research Unit and UNWTO Monitoring Centre for the Waikato. He has over 200 refereed journal articles and 14 books with his name. He has undertaken reports for UNWTO in Xishuangbanna Autonomous Prefecture and in Changsha, China, and been a members of various committees associated with tourism and New Zealand's governmental ministries. He completed his doctoral degree at the Aston University Management School in Birmingham and holds degrees from Notting Trent University, Nottingham University and London University.