

Why Is The Partial Least Square Important For Tourism Studies

Sari Lenggogeni^{*1}

¹Senior Lecturer in Faculty of Economics, Universitas Andalas, Indonesia,
West Sumatra, Indonesia

²Director of Tourism Development Centre, Universitas Andalas, Indonesia,
West Sumatra, Indonesia

sarilenggogeni@eb.unand.ac.id

Abstract

Although the multiple regression method has been applied to exploratory research on most tourism studies, there is lack of understanding on studies that present a well-justified rationale in choosing a robust statistical tool for data analysis. This research note aims to review why tourism researchers are encouraged to use the Partial Least Squares Structural Equation Modelling (PLS-SEM) method to address this research problem. This article provides rationale, comparisons among techniques for multiple regression-based papers and suggestions for tourism researchers to justify why PLS-SEM is important for exploratory studies.

Keywords: structural equation modelling, partial least square, tourism method, multiple-regression



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Introduction

A considerable amount of literature using the PLS-SEM method has been published in top rank tourism journals, such as the Journal of Travel Research, Tourism Management and Tourism Economics (Sarstedt et al., 2020). This trend aligns with the growth of quantitative research publications based on multiple regression analysis in the tourism discipline. Whilst the PLS-SEM has been mainly considered as a tool for the complex-inter relationship between endogenous and exogenous variables in the research framework model, there are many papers that have not presented a strong justification on why they should choose the PLS-SEM method.

Understanding the nature, problem and method in research, including its data analytical tool, is a fundamental key in the robustness of research. Selecting an inaccurate method for research could result in misleading interpretations, findings and analyses. This is most likely true for most quantitative research studies that adapted a multiple regression analysis method, particularly in tourism studies. It is important to gain a better insight on the nature of research, rationale, and

justification in order to choose a structural equation modelling for a complex research framework model in the tourism discipline.

Structural Equation Modelling in Tourism Research and Theory Development

Most tourism studies are exploratory-based research papers in nature (i.e., Devianto, Ridho, Maryati, & Lenggogeni, 2019; Filimonau & Mika, 2019; Mason, Augustyn, & Seakhwa-King, 2010; Nunkoo & Ramkissoon, 2011). There are many development research theories, such as a new scale development in tourism study (for instance, Lenggogeni, Ritchie, & Slaughter, 2019), which is often applied in the complex multiple regression-based research framework (Lenggogeni, 2015; Lenggogeni & Saito, 2018).

There are different approaches to regression-based research models that might considerably be useful for tourism researchers in addressing research problems. They are categorized into first generation (multiple regression) and second generation (Structural Equation Modelling) models (Lowry & Gaskin, 2014; Muthén, 2001). These approaches are frequently used in research

on causal relationships and offer some advantages and disadvantages for each analysis. Nevertheless, what remains unclear is most of these studies did not meet the rule of thumbs, appropriateness, justification as well as rationale to justify their data analysis.

Furthermore, one should bear in mind that tourism, travel and hospitality research tend to have more error/ bias results. For instance, a respondent's perceptions and opinions, which are mostly found in research based on a respondent's perceptions (Yüksel, 2017). Error or biased results could also occur due to inaccurate responses from respondents. Therefore, if a tourism researcher intends to propose a research framework model that is highly complex or contains multiple equations in the path analysis to investigate the multiple relationships between latent variables or manifest variables, the Structural Equation Modelling (SEM) model is the most suitable option.

The Structural Equation Modelling (SEM) model offers the advantage of minimizing large errors in each equation and simplifying the analysis with a large number of construct use in one research validating the instrument while running the analysis simultaneously (Henseler, Ringle, & Sinkovics, 2009; Nunkoo & Ramkissoon, 2011). Chin (1998, p. vii) claimed that the SEM is an advantage for the first generation of multivariate analysis with multiple regressions because (a) the relationship model uses multiple predictors and criterion variables, (b) it may employ unobservable Latent Variables (LVs), (c) there is a model for errors in measurements for observed variables, and (d) statistically tests a priori substantive/theoretical and measurement assumption against empirical data (i.e., confirmatory analysis). However, to prevent misleading conclusions and gain a valid insight prior to the utilization of the SEM model as a statistical analysis tool, the researcher must first take into account theoretical, methodological and statistical analyses as three assessment indicators for the SEM method (Bagozzi, 1981, p. 375). Hair et al., (2014) argued that the researcher must choose approaches based on the context of the research as well as the suitability of the goal of the research. Therefore, the researcher should understand the characteristics of each of the tools and the objectives of these two different approaches; Co-Variance-based SEM and Variance-based (Component) SEM.

Co-Variance-based SEM and Variance-based (Component-based) SEM

There are basically two types of approaches to key research methods for estimating the relationship in a structural equation model, which are Co-Variance Based SEM (CB-SEM) using AMOS, LISREL and EQS packages and Variance-based (Component-based) SEM or PLS-SEM using Partial least Square packages (Haenlein & Kaplan, 2004, p. 285; Henseler et al., 2009; Hair et al., 2014).

Both of these SEM tools are widely applied in social science research, such as marketing research, because these two groups of "SEM methods are robust for estimating causal models with latent variables and provide simultaneous equation with a measurement for errors" (Henseler et al., 2009, p. 310). The CB-SEM method, in particular, gained more attention from researchers in terms of its application for multi relationship variable research and is claimed to be the most popular one between the two methods (Haenlein & Kaplan, 2004). The CB-SEM method provides the most efficient parameter estimates and overall test of model fits (Anderson & Gerbing, 1988). The application of Variance-based SEM like PLS has been gaining popularity in the last five years as it has grown acceptance in high reputable journals in the marketing and strategic management discipline (Hair et al., 2014). This tool's second generation technique is widely applied in social science research in multivariate analysis (Hair et al., 2014). What is interesting is that Variance-based SEM is sometimes used as an alternative method for the CB-SEM method if the various assumptions required in the CB-SEM method is unsatisfactory as a proposal for one's research (Henseler et al., 2009; Monecke & Leisch, 2012). This approach is claimed as the most powerful analysis tool compared to the CB-SEM method (Henseler et al., 2009, p. 77). However, the results from using the CB-SEM or PLS-SEM methods do not differ much. Compared to the rule of thumb for the CB-SEM method related to the "normality of distributions, minimum sample size, and maximum model complexity, or related methodological anomalies that occur in the process of model estimation", the PLS-SEM method is considered as an alternative approach for theory testing (Hair et al p. 18). The PLS-SEM method is not sensitive to the case of small samples like the CB-SEM method and it works with a large number of constructs and indicators.

Hair et al (2014, p.23) postulated that when it comes to larger data sets ($n = 250 +$) and larger number of indicators on variables (4+), the CB-SEM and PLS-SEM have similar results. In addition, applications of metric, quasi-metric, and categorical data are allowed in PLS-SEM (Hair et al., 2014). Furthermore, Fornell and Larcker (1981) suggested that once the model has no explanatory power, although their nature of research is theory-testing oriented, the researcher needs to focus on the relationship between observable constructs.

CB SEM and PLS SEM Goals

Based on goals, the Covariance-based SEM (CB-SEM) method is primarily employed to confirm (or reject) theories, or “theory-testing”, while the Variance-based SEM or PLS –SEM method is primarily utilized to develop theories, or “theory-building”, in exploratory research or in causal modelling application (Anderson & Gerbing, 1988; Henseler et al., 2009; Joseph F Hair et al., 2014, p. 4). The CB-SEM method is aimed at using the model to explain the co-variation among all indicators. On the other hand, the PLS path modelling maximizes the explained variance of all dependent variables. Therefore, this supports a prediction-oriented goal instead of confirming a theory (Henseler et al., 2009). Hair (2011, p.18) added that compared to the CB-SEM method, the PLS-SEM method is appropriately used when there is little prior knowledge on structural model relationships or the measurement of the construct or when the emphasis is more on exploration than confirmation. In other words, the PLS- SEM method is suitable if the primary objective of applying structural modelling is the prediction and explanation of a target construct where the theory is less developed. The PLS- SEM method is appropriate when the research objective is theory development and explanation of variance (prediction of the construct) (Anderson & Gerbing, 1988; Joseph F Hair et al., 2014, p. 14).

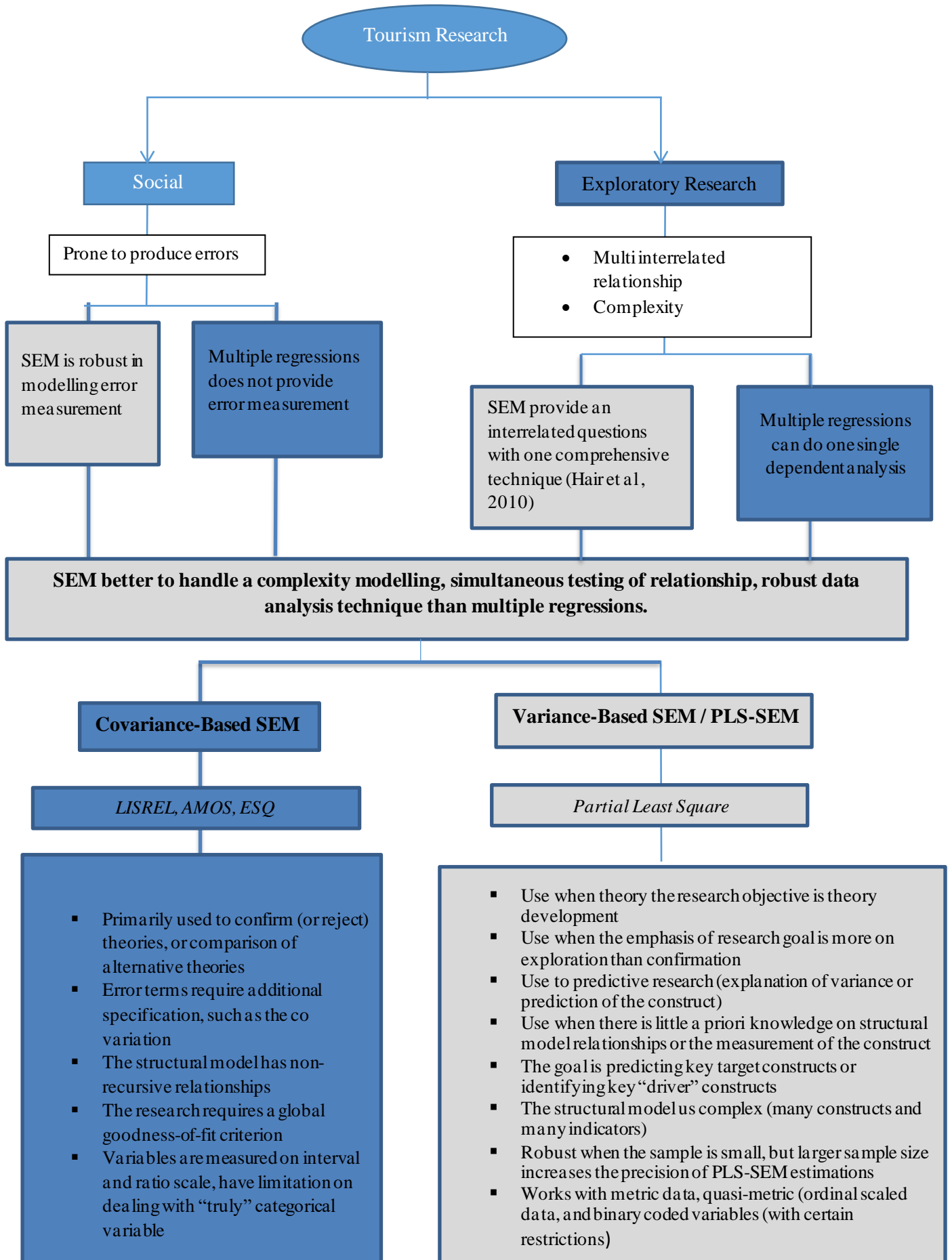
One interesting note is the CB-SEM can also be used for either theory testing or theory building. Haenlin & Kaplan highlighted that “SEM can be (and often is) used to test (and consequently to either support or reject) theoretical assumptions with empirical data” (2004, p. 286). In the case for research aimed at theory testing, the researcher

needs to align with an adequate theoretical concept, as Fornell and Larcker (1981) argued that the CB-SEM is sensitive to a lack of significant theory. Hair et al (2011, p.620) postulated, “In all instance SEM analyses should be dictated first and foremost by a strong theoretical base”. It means that the CB-SEM method is strongly theoretical based. Haenlin & Kaplan (2004, p.286) strengthened this statement by saying that the CB-SEM method is often used as theory testing (either support or reject) theoretical assumption with empirical data. However, the SEM method is also allowed to be used for theory development and is commonly used for predictive-oriented research like PLS-SEM. Therefore, before deciding to use this statistical tool, the researcher must prepare a strong theoretical concept as a rule of thumb.

Meanwhile, it is important to understand about theory development. Theory is defined as “that body of logically interconnected propositions which provides an interpretative basis for understanding phenomenon” (Dann, Nash, & Pearce, 1988, p. 4). Bagozzi and Phillips (1982, p. 465) claimed that there are three different types of concepts for theories:

a) theoretical concept “unobservable properties or attributes of a social unit or entity. “b) empirical concept “properties or relations whose presence or absence in a given case can be inter subjectively ascertained, under suitable circumstances, by direct observation” c) derived concept “like theoretical concepts, are unobservable. Unlike theoretical concepts, however, derived concepts must be tied directly to empirical concepts, and they are typically at lower levels of abstraction than theoretical concepts”. The relationship built in the researcher’s research context here refers to the third type of Bagozzi’s concept, which are the “correspondence rules” that “link to **theoretical or derived to empirical concept** and serve to provide empirical significance to theoretical terms” (1984. p.17). Hence, the researcher needs to understand the underlying theory and purpose of his or her research that leads to the understanding of theory testing or theory building in order to choose the appropriate statistical analysis tool (Henseler et al., 2009).

Figure 1: SEM in Tourism Research



Source : Developed for this research (Hair, Black, Babin & Anderson(2011) ; Hair, Hult, Ringle & Sarstedt(2014) ; Haenlin and Kaplan (2004); Nunkoo & Ramkinsoon(2011) ; Henseler, Ringle & Sinkovics(2009)

Figure 1 provides an explanation on the framework for a multiple regression-based hierarchy rationale applied in tourism studies

Rationale for Choosing PLS-SEM

In this case, we will use an example of exploratory and predictive research with the following characteristics: interrelated relationship (multiple equations), complexity of model, predictive model and theory development. While there are three options of statistical analysis tools that may help with this research (multiple regression, CB-SEM and PLS- SEM), the four arguments below may help to justify that there is only one appropriate tool for this research sample.

First, the context of the study claims to be predictive rather than confirmatory. It also is also a model with high complexity. The PLS-SEM method is confirmed to be the most suitable methodological analysis than the CB-SEM or ML. As this study emphasizes on causal relationships and theory development using Exploratory Factor Analysis, the PLS-SEM method is the most appropriate method for this research's questions. Why? Because the PLS-SEM is more for suitable for research that has a predictive goal instead of a confirmatory goal. Compared to Covariance-Based SEM methods, like LISREL or AMOS which use Maximum Likelihood (ML), the PLS-SEM method is more often used for predictive-oriented research. LISREL founder Jöreskog (1982, p. 270) postulated that "ML is theory-oriented, emphasizes the transition from an exploratory to confirmatory factor analysis. PLS is primarily intended for causal-predictive analysis in situations of high complexity but low theoretical information".

Second, the study is part of a new theory development that uses an exploratory factor analysis to find a new scale. This new scale was developed based on the theoretical concept of one theory on the empirical-based evidence. Thus, "the derived concept" is a result of empirical significance to theoretical terms, in which this relationship is defined as correspondence rules (Bagozzi, 1984, p. 17). The PLS-SEM is an appropriate tool because the researcher aims to test a new scale in a multiple regression-based

analysis that has not been tested in previous studies. Hair et al (2012) suggested that the CB-SEM method is the opposite of an exploratory technique. The researcher needs to specify which variables are associated with each construct and which variables require a theoretical base to explain each relationship path. Therefore, employing confirmatory factor analysis is not plausible because it cannot be applied to strong existing theoretical concepts. This also true on how the path analysis was built upon this new scale - without a strong former theory - which lead the researcher to suggest choosing the PLS-SEM method rather than the CB-SEM method (i.e AMOS, Lisrel). Ainuddin, Beamish, Hulland and Rouse (2007) stated that the "use of PLS is especially suited to exploratory studies such as this, where the measures [...] are new and the relationships [...] have not been previously tested" (p.56). Meanwhile, Tsang (2002, p. 841) argued that "PLS is particularly suitable for data analysis during the early stages of theory development where the theoretical model and its measure are not well formed".

This concludes that for theory development research, the PLS-SEM method is more suitable for a regression-based approach in the second-generation compared to CB-SEM, as supported by Venaik, Midgley and Devinney (2005, p. 665), who stated, "At an early stage of development [...] the regression based approach of PLS is considered more appropriate than covariance-based methods such as LISREL.". Fornell & Lacker (1981) also supported this argument by postulating that PLS is one of the second-generation of multivariate analysis approaches (LISREL, AMOS and EQS are others) that combine theoretical and empirical knowledge in order to maximize the variance explained. Henseler et al., (2009) also asserted that CB-SEM method is the most appropriate statistical methodology in modelling where the prior theory is strong and further testing development is the goal (p.296). In addition, Anderson & Gerbing (1988) noted that the CB SEM method is suitable to "theory-testing" rather than "theory building". Based on these arguments concerning the choice between Co-Variance and Variance-Based SEM

models, tourism research with a theory development goal is suggested to opt to use the PLS-SEM method because it is more suitable based on the second criteria for PLS-SEM justification.

Third, this research uses a non-metric scale measurement. If more than one categorical latent variable is applied in a research, then using the PLS-SEM method as a statistical analysis tool is more appropriate than the CB-SEM method, like AMOS or Lisrel. This was also justified by Bagozzi (1984, p. 384), who postulated that the LISREL's disadvantage is that it was designed for variables measured based on at least interval or ordinal scales, as suggested that "if many scale steps are used (say five to seven or more) LISREL may be used for most". Meanwhile, this study uses nominal data using the dichotomous scale, which means the PLS-SEM is a more suitable methodological analysis tool. Mintu-Wimsat and Graham (2004, p. 352) use the PLS-SEM method because the PLS is able to minimize biases associated with [...] dichotomous and ordinal measures". Falk & Miller (1992; 5-6 in Gracia, 2007) also supported this argument by stating PLS is flexible when a researcher's manifest variables

are categorical or if they have different levels of measurement".

Fourth, the complex model and the large number of constructs. If a tourism research paper is proposing a highly complex model and involves multiple constructs that consist of large number of indicators, the PLS-SEM method is more appropriate as a data analysis tool. Hair et al (2014); Haenlin and Kaplan (2004), agreed that there are no limitations on highly complex models when using the PLS-SEM method, which also allows a large number of indicators. This argument was also supported by Henseler et al (2009, p.283), who claimed that PLS could estimate highly complex models with latent and manifest variables. Likewise, Anderson and Gerbing Anderson and Gerbing (1984) argued CB-SEM is not suitable for more complex research models because it could be affected by the decline of goodness-of-fit-indices.

Overall, Table 1 provides a summary of the advantages and disadvantages of Multiple Regression, SEM with Covariance Bases (AMOS, LISREL, EQS), SEM with Component Based (Partial Least Square).

Table 1: Comparison between First- and Second-Generation Regression Models

Multiple regression	Covariance-Based SEM	PLS-SEM	
Tests regression models with one only dependent variable at a time	Able to answer complex research questions and test multivariate models		
Ignores the modelling of measurement errors	SEM takes measurement errors into consideration		
Multiple regression cannot propose new relationships in a model that is theoretically justified while at the same time fits with the model's data	Confirms or tests theories	Theory-building	
Limited to testing a model that contains a single dependent variable.	SEM has the ability to test a full structural model simultaneously—reduces risk of lack connection		
	Explanatory research	Exploratory research	
	Requires large sample	Not sensitive to small or large samples	
		No issues in analysing formative and reflective indicators	
		Complex models	
		Allows the use of a large number of constructs and indicators	

Source: Developed from Nunkoo & Ramkinsoon (2011); Hair et al (2014); Henseler (2009); Anderson and Gerbing (1982)

Conclusion

The above discussion highlights the importance of presenting a strong justification for choosing the appropriate data statistical analysis tool for regression-based research in the first and second generation. In order to avoid the fundamental mistake often found in research methods, which is misleading data interpretation and analysis, this article has unravelled a comparison of each

approach in order to choose the most suitable type of tool for research along with each approach's advantages and disadvantages. Future exploratory research in tourism studies is suggested to in to carefully understand the nature and aim of the research in addition to presenting a rigorous method of justification that could produce a sound and robust research paper.

References

Ainuddin, R. A., Beamish, P. W., Hulland, J. S., & Rouse, M. J. (2007). Resource attributes and firm performance in international joint ventures. *Journal of World Business, 42*(1), 47-60. doi:http://dx.doi.org/10.1016/j.jwb.2006.11.001

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*(2), 155-173. doi:10.1007/BF02294170

- Anderson, J. C., & Gerbing, D. W. (1988). Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin*, 103(3), 411-423. Retrieved from http://www.cob.unt.edu/slides/Paswan/BUSI6280/Anderson_Gerbing_1988.pdf
- Bagozzi, R. P. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error: A Comment. *Journal of Marketing Research*, 18(3), 375-381. doi:10.2307/3150979
- Bagozzi, R. P. (1984). A Prospectus for Theory Construction in Marketing. *Journal of Marketing*, 48(1), 11-29. doi:10.2307/1251307
- Bagozzi, R. P., & Phillips, L. W. (1982). Representing and Testing Organizational Theories: A Holistic Construal. *Administrative Science Quarterly*, 27(3), 459-489. Retrieved from <http://www.jstor.org/stable/2392322>
- Chin, W. W. (1998). Commentary: Issues and Opinion on Structural Equation Modeling. *MIS Quarterly*, 22(1), vii-xvi. doi:10.2307/249674
- Dann, G., Nash, D., & Pearce, P. (1988). Methodology in tourism research. *Annals of Tourism Research*, 15(1), 1-28. doi:[http://dx.doi.org/10.1016/0160-7383\(88\)90068-0](http://dx.doi.org/10.1016/0160-7383(88)90068-0)
- Devianto, D., Ridho, M., Maryati, S., & Lenggogeni, S. (2019). *Path analysis of entrepreneurial motivations in tourism based on local resources and creative economy in nagari salayo of west sumatra*. Paper presented at the Journal of Physics: Conference Series.
- Filimonau, V., & Mika, M. (2019). Return labour migration: an exploratory study of Polish migrant workers from the UK hospitality industry. *Current Issues in Tourism*, 22(3), 357-378.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. doi:10.2307/3151312
- Gil-Gracia, J. R. (2007). *Using Partial Least Squares (PLS) for Digital Government Research*. Paper presented at the Annual International Digital Government Research Conferences, Montreal-Canada.
- Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding Statistics*, 3(4), 283-297. doi:10.1207/s15328031us0304_4
- Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2013). *A primer on partial least squares structural equation modeling (PLS-SEM)*: SAGE Publications, Incorporated.
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modelling in international marketing. *Advances in International Marketing*, 20, 277-319. doi:[http://dx.doi.org/10.1108/S1474-7979\(2009\)0000020014](http://dx.doi.org/10.1108/S1474-7979(2009)0000020014)
- Jöreskog, K. G. (1982). The LISREL approach to causal model-building in the social sciences. *Systems under indirect observation, Part I*, 81-100.
- Joseph F Hair, J., Black, W. C., Babin, B. J., Anderson, R. E., & L.Tatham, R. (2011). *Multivariate Data Analysis* (7th ed.). New Jersey: Pearson Prentice Hall.
- Joseph F Hair, J., Hult, G. T. M., Ringle, C. M., & Sasrstedt, M. (2014). *A Primer On Partial Least Squares Structural Equation Modelling (PLS-SEM)*. Los Angeles, London, New Delhi, Singapore, Washington DC: SAGE Publication, Inc.
- Lenggogeni, S. (2015). *Travel risk perceptions, travel intentions and influencing factors: A natural disaster context*. (Doctoral Dissertation). The University of Queensland, Australia.
- Lenggogeni, S., Ritchie, B. W., & Slaughter, L. (2019). Understanding travel risks in a developing country: a bottom up approach. *Journal of Travel & Tourism Marketing*, 36(8), 941-955.
- Lenggogeni, S., & Saito, H. (2018). Does a religion matter in travel risk and behavioural intention in earthquake vulnerable cities?: A case for domestic traveller in Indonesia and Japan. *CAUTHE 2018: Get Smart: Paradoxes and Possibilities in Tourism, Hospitality and Events Education and Research*, 836.
- Lowry, P. B., & Gaskin, J. (2014). Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *IEEE transactions on professional communication*, 57(2), 123-146.
- Mason, P., Augustyn, M., & Seakhoa-King, A. (2010). Exploratory study in tourism: Designing an initial, qualitative phase of sequenced, mixed methods research. *International Journal of Tourism Research*, 12(5), 432-448.

- Mintu-Wimsatt, A., & Graham, J. L. (2004). Testing a Negotiation Model on Canadian Anglophone and Mexican Exporters. *Journal of the Academy of Marketing Science*, 32(3), 345-356. doi:10.1177/0092070304266123
- Monecke, A., & Leisch, F. (2012). semPLS: Structural Equation Modeling Using Partial Least Squares. *Journal of Statistical Software*, 48(3), 1-32. Retrieved from <http://www.jstatsoft.org/v48/i03/paper>
- Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class–latent growth modeling.
- Nunkoo, R., & Ramkissoon, H. (2011). Structural equation modelling and regression analysis in tourism research. *Current Issues in Tourism*, 1-26. doi:10.1080/13683500.2011.641947
- Sarstedt, M., Ringle, C. M., Cheah, J.-H., Ting, H., Moisescu, O. I., & Radomir, L. (2020). Structural model robustness checks in PLS-SEM. *Tourism Economics*, 26(4), 531-554.
- Tsang, E. W. K. (2002). Acquiring knowledge by foreign partners from international joint ventures in a transition economy: learning-by-doing and learning myopia. *Strategic Management Journal*, 23(9), 835-854. doi:10.1002/smj.251
- Venaik, S., Midgley, D. F., & Devinney, T. M. (2005). Dual paths to performance: the impact of global pressures on MNC subsidiary conduct and performance. *J Int Bus Stud*, 36(6), 655-675. Retrieved from <http://dx.doi.org/10.1057/palgrave.jibs.8400164>
- Yüksel, A. (2017). A critique of “Response Bias” in the tourism, travel and hospitality research. *Tourism management*, 59, 376-384.