Received 27 November 2018
Revised 30 November 2018
Accepted 1 December 2018

When to use and how to report the results of PLS-SEM

Joseph F. Hair
Department of Marketing and Quantitative Methods, University of South Alabama, Mobile, Alabama, USA

Jeffrey J. Risher
Department of Marketing, Supply Chain Logistics and Economics, University of West Florida College of Business, Pensacola, Florida, USA and Department of Business, University of Mobile, Mobile, Alabama, USA

Marko Sarstedt
Faculty of Economics and Management, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany and Monash University of Malaysia, Bandar Sunway, Malaysia, and

Christian M. Ringle
Institute of Human Resource Management and Organizations, Hamburg University of Technology (TUHH), Hamburg, Germany and the University of Waikato, Hamilton, New Zealand

Abstract

Purpose – The purpose of this paper is to provide a comprehensive, yet concise, overview of the considerations and metrics required for partial least squares structural equation modeling (PLS-SEM) analysis and result reporting. Preliminary considerations are summarized first, including reasons for choosing PLS-SEM, recommended sample size in selected contexts, distributional assumptions, use of secondary data, statistical power and the need for goodness-of-fit testing. Next, the metrics as well as the rules of thumb that should be applied to assess the PLS-SEM results are covered. Besides presenting established PLS-SEM evaluation criteria, the overview includes the following new guidelines: PLSpredict (i.e., a novel approach for assessing a model's out-of-sample prediction), metrics for model comparisons, and several complementary methods for checking the results' robustness.

Design/methodology/approach – This paper provides an overview of previously and recently proposed metrics as well as rules of thumb for evaluating the research results based on the application of PLS-SEM.

Findings – Most of the previously applied metrics for evaluating PLS-SEM results are still relevant. Nevertheless, scholars need to be knowledgeable about recently proposed metrics (e.g. model comparison criteria) and methods (e.g. endogeneity assessment, latent class analysis and PLSpredict), and when and how to apply them to extend their analyses.

Research limitations/implications – Methodological developments associated with PLS-SEM are rapidly emerging. The metrics reported in this paper are useful for current applications, but must always be up to date with the latest developments in the PLS-SEM method.

Originality/value – In light of more recent research and methodological developments in the PLS-SEM domain, guidelines for the method's use need to be continuously extended and updated. This paper is the

European Business Review
Vol. 31 No. 1, 2019
pp. 2-24

0955-534X
DOI 10.1108/EBR-11-2018-0203

Even though this article does not use the statistical software SmartPLS (www.smartpls.com), Ringle acknowledges a financial interest in SmartPLS.

most current and comprehensive summary of the PLS-SEM method and the metrics applied to assess its solutions.

Keywords Structural equation modeling, Partial least squares, PLS-SEM, Model comparisons, PLSpredict

Paper type General review

Introduction

For many years, covariance-based structural equation modeling (CB-SEM) was the dominant method for analyzing complex interrelationships between observed and latent variables. In fact, until around 2010, there were far more articles published in social science journals that used CB-SEM instead of partial least squares structural equation modeling (PLS-SEM). In recent years, the number of published articles using PLS-SEM increased significantly relative to CB-SEM (Hair *et al.*, 2017b). In fact, PLS-SEM is now widely applied in many social science disciplines, including organizational management (Sosik *et al.*, 2009), international management (Richter *et al.*, 2015), human resource management (Ringle *et al.*, 2019), management information systems (Ringle *et al.*, 2012), operations management (Peng and Lai, 2012), marketing management (Hair *et al.*, 2012b), management accounting (Nitzl, 2016), strategic management (Hair *et al.*, 2012a), hospitality management (Ali *et al.*, 2018b) and supply chain management (Kaufmann and Gaeckler, 2015). Several textbooks (e.g., Garson, 2016; Ramayah *et al.*, 2016), edited volumes (e.g., Avkiran and Ringle, 2018; Ali *et al.*, 2018a), and special issues of scholarly journals (e.g., Rasoolimanesh and Ali, 2018; Shiau *et al.*, 2019) illustrate PLS-SEM or propose methodological extensions.

The PLS-SEM method is very appealing to many researchers as it enables them to estimate complex models with many constructs, indicator variables and structural paths without imposing distributional assumptions on the data. More importantly, however, PLS-SEM is a causal-predictive approach to SEM that emphasizes prediction in estimating statistical models, whose structures are designed to provide causal explanations (Wold, 1982; Sarstedt *et al.*, 2017a). The technique thereby overcomes the apparent dichotomy between explanation – as typically emphasized in academic research – and prediction, which is the basis for developing managerial implications (Hair *et al.*, 2019). Additionally, user-friendly software packages are available that generally require little technical knowledge about the method, such as PLS-Graph (Chin, 2003) and SmartPLS (Ringle *et al.*, 2015; Ringle *et al.*, 2005), while more complex packages for statistical computing software environments, such as R, can also execute PLS-SEM (e.g. semPLS; Monecke and Leisch, 2012). Authors such as Richter *et al.* (2016), Rigdon (2016) and Sarstedt *et al.* (2017a) provide more detailed arguments and discussions on when to use and not to use PLS-SEM.

The objective of this paper is to explain the procedures and metrics that are applied by editors and journal review boards to assess the reporting quality of PLS-SEM findings. We first summarize several initial considerations when choosing to use PLS-SEM and cover aspects such as sample sizes, distributional assumptions and goodness-of-fit testing. Then, we discuss model evaluation, including rules of thumb and introduce important advanced options that can be used. Our discussion also covers PLSpredict, a new method for assessing a model's out-of-sample predictive power (Shmueli *et al.*, 2016; Shmueli *et al.*, 2019), which researchers should routinely apply, especially when drawing conclusions that affect business practices and have managerial implications. Next, we introduce several complementary methods for assessing the results' robustness when it comes to measurement model specification, nonlinear structural model effects, endogeneity and unobserved heterogeneity (Hair *et al.*, 2018; Latan, 2018). Figure 1 illustrates the various aspects that we discuss in the following sections.

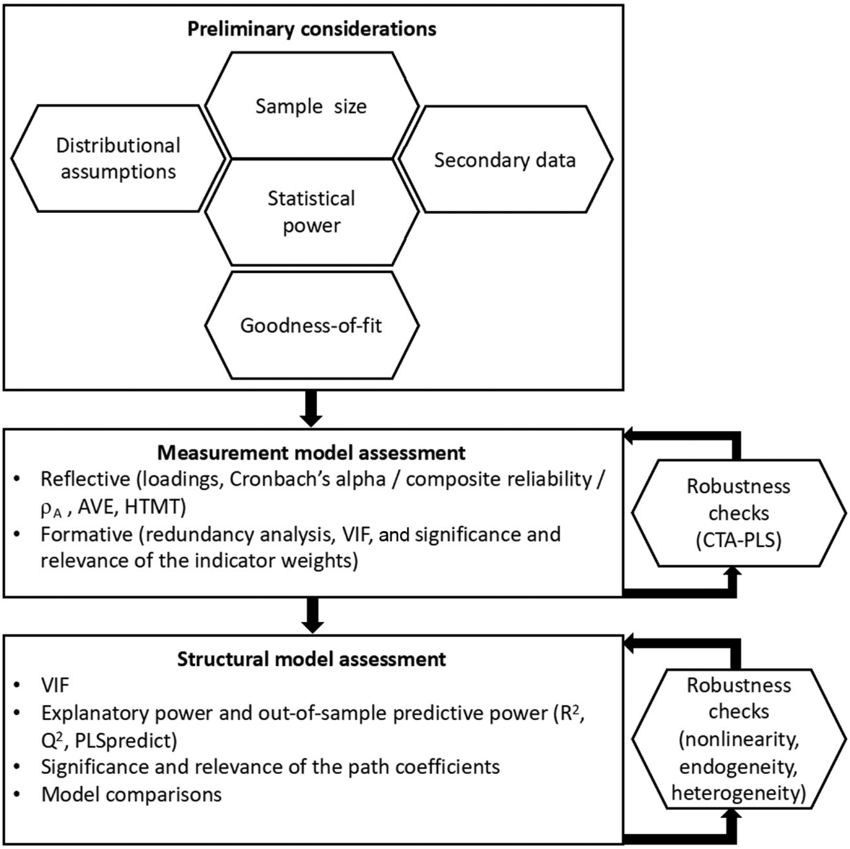

Figure 1.
Aspects and statistics to consider in a PLS-SEM analysis

Preliminary considerations

The Swedish econometrician Herman O. A. Wold (1975, 1982, 1985) developed the statistical underpinnings of PLS-SEM. The method was initially known and is sometimes still referred to as PLS path modeling (Hair *et al.*, 2011). PLS-SEM estimates partial model structures by combining principal components analysis with ordinary least squares regressions (Mateos-Aparicio, 2011). This method is typically viewed as an alternative to Jöreskog's (1973) CB-SEM, which has numerous – typically very restrictive – assumptions (Hair *et al.*, 2011).

Jöreskog's (1973) CB-SEM, which is often executed by software packages such as LISREL or AMOS, uses the covariance matrix of the data and estimates the model parameters by only considering common variance. In contrast, PLS-SEM is referred to as variance-based, as it accounts for the total variance and uses the total variance to estimate parameters (Hair *et al.*, 2017b).

In the past decade, there has been a considerable debate about which situations are more or less appropriate for using PLS-SEM (Goodhue *et al.*, 2012; Marcoulides *et al.*, 2012; Marcoulides and Saunders, 2006; Rigdon, 2014a; Henseler *et al.*, 2014; Khan *et al.*, 2019). In the following sections, we summarize several initial considerations when to use PLS-SEM (Hair *et al.*, 2013). Furthermore, we compare the differences between CB-SEM and PLS-SEM (Marcoulides and Chin, 2013; Rigdon, 2016). In doing so, we note that recent research has moved beyond the CB-SEM versus PLS-SEM debate (Rigdon *et al.*, 2017; Rigdon, 2012), by establishing PLS-SEM as a distinct method for analyzing composite-based path models. Nevertheless, applied research is still confronted with the choice between the two SEM methods. Researchers should select PLS-SEM:

- when the analysis is concerned with testing a theoretical framework from a prediction perspective;
- when the structural model is complex and includes many constructs, indicators and/or model relationships;
- when the research objective is to better understand increasing complexity by exploring theoretical extensions of established theories (exploratory research for theory development);
- when the path model includes one or more formatively measured constructs;
- when the research consists of financial ratios or similar types of data artifacts;
- when the research is based on secondary/archival data, which may lack a comprehensive substantiation on the grounds of measurement theory;
- when a small population restricts the sample size (e.g. business-to-business research); but PLS-SEM also works very well with large sample sizes;
- when distribution issues are a concern, such as lack of normality; and
- when research requires latent variable scores for follow-up analyses.

The above list provides an overview of points to consider when deciding whether PLS is an appropriate SEM method for a study.

Sample size

PLS-SEM offers solutions with small sample sizes when models comprise many constructs and a large number of items (Fornell and Bookstein, 1982; Willaby *et al.*, 2015; Hair *et al.*, 2017b). Technically, the PLS-SEM algorithm makes this possible by computing measurement and structural model relationships separately instead of simultaneously. In short, as its name implies, the algorithm computes partial regression relationships in the measurement and structural models by using separate ordinary least squares regressions. Reinartz *et al.* (2009), Henseler *et al.* (2014) and Sarstedt *et al.* (2016b) summarize how PLS-SEM provides solutions when methods such as CB-SEM develop inadmissible results or do not converge with complex models and small sample sizes, regardless of whether the data originates from a common or composite model population. Hair *et al.* (2013) indicate that certain scholars have falsely and misleadingly taken advantage of these characteristics to generate solutions with extremely small sample sizes, even when the population is large and accessible without much effort. This practice has unfortunately damaged the reputation of PLS-SEM to some extent (Marcoulides *et al.*, 2009). Like other multivariate methods, PLS-SEM is not capable of turning a poor (e.g. non-representative) sample into a proper one to obtain valid model estimations.

PLS-SEM can certainly be used with smaller samples but the population's nature determines the situations in which small sample sizes are acceptable (Rigdon, 2016). Assuming that other situational characteristics are equal, the more heterogeneous the population, the larger the sample size needed to achieve an acceptable sampling error (Cochran, 1977). If basic sampling theory guidelines are not considered (Sarstedt *et al.*, 2018), questionable results are produced. To determine the required sample size, researchers should rely on power analyses that consider the model structure, the anticipated significance level and the expected effect sizes (Marcoulides and Chin, 2013). Alternatively, Hair *et al.* (2017a) have documented power tables indicating the required sample sizes for a variety of measurement and structural model characteristics. Finally, Kock and Hadaya (2018) suggest the inverse square root method and the gamma-exponential method as two new approaches for minimum sample size calculations.

Akter *et al.* (2017) note that most prior research on sample size requirements in PLS-SEM overlooked the fact that the method also proves valuable for analyzing large data quantities. In fact, PLS-SEM offers substantial potential for analyzing large data sets, including secondary data, which often does not include comprehensive substantiation on the grounds of measurement theory (Rigdon, 2013).

Distributional assumptions

Many scholars indicate that the absence of distributional assumptions is the main reason for choosing PLS-SEM (Hair *et al.*, 2012b; Nitzl, 2016; do Valle and Assaker, 2016). While this is clearly an advantage of using PLS-SEM in social science studies, which almost always rely on nonnormal data, on its own, it is not a sufficient justification.

Scholars have noted that maximum likelihood estimation with CB-SEM is robust against violations of normality (Chou *et al.*, 1991; Olsson *et al.*, 2000), although it may require much larger sample sizes (Boomsma and Hoogland, 2001). If the size of the data set is limited, CB-SEM can produce abnormal results when data are nonnormal (Reinartz *et al.*, 2009), while PLS-SEM shows a higher robustness in these situations (Sarstedt *et al.*, 2016b).

It is noteworthy that in a limited number of situations, nonnormal data can also affect PLS-SEM results (Sarstedt *et al.*, 2017a). For instance, bootstrapping with nonnormal data can produce peaked and skewed distributions. The use of the bias-corrected and accelerated (BCa) bootstrapping routine handles this issue to some extent, as it adjusts the confidence intervals for skewness (Efron, 1987). Only choosing PLS-SEM for data distribution reasons is, therefore, in most instances not sufficient, but it is definitely an advantage in combination with other reasons for using PLS-SEM.

Secondary data

Secondary (or archival) data are increasingly available to explore real-world phenomena (Avkiran and Ringle, 2018). Research which is based on secondary data typically focuses on a different objective than in a standard CB-SEM analysis, which is strictly confirmatory in nature. More precisely, secondary data are mainly used in exploratory research to propose causal relationships in situations which have little clearly defined theory (Hair *et al.*, 2017a, 2017b). Such settings require researchers to put greater emphasis on examining all possible relationships rather than achieving model fit (Nitzl, 2016). By its nature, this process creates large complex models that cannot be analyzed with the full information CB-SEM method. In contrast, the iterative approach of PLS-SEM uses limited information, making the method more robust and not constrained by the requirements of CB-SEM (Hair *et al.*, 2014). Thus, PLS-SEM is very suitable for exploratory research with secondary data, because it offers the flexibility needed for the interplay between theory and data (Nitzl, 2016) or, as Wold (1982 p. 29) notes, "soft modeling is primarily designed for research contexts that are simultaneously data-rich and theory-skeletal." Furthermore, the increasing popularity of secondary data analysis (e.g. by using data that stem from company databases, social media, customer tracking, national statistical bureaus or publicly available survey data) shifts the research focus from strictly confirmatory to predictive and causal-predictive modeling. Such research settings are a perfect fit for the prediction-oriented PLS-SEM approach.

PLS-SEM also proves valuable for analyzing secondary data from a measurement theory perspective. Unlike survey measures, which are usually crafted to confirm a well-developed theory, measures used in secondary data sources are typically not created and refined over time for confirmatory analyses (Sarstedt and Mooi, 2019). Thus, achieving model fit with secondary data measures is unlikely in most research situations when using CB-SEM. Furthermore, researchers who use secondary data do not have the opportunity to revise or refine the measurement model to achieve fit. Another major advantage of PLS-SEM in this context is that it permits the unrestricted use of single-item and formative measures (Hair *et al.*, 2014). This is extremely valuable for archival research, because many measures are actually artifacts found in corporate databases, such as financial ratios and other firm-fixed factors (Richter *et al.*, 2016). Often, several types of financial data may be used to create an index as a measure of performance (Sarstedt *et al.*, 2017a, 2017b). For instance, Ittner *et al.* (1997) operationalized strategy with four indicators as follows: the ratio of research and development to sales, the market-to-book ratio, the ratio of employees to sales and the number of new product or service introductions. Similarly, secondary data could be used to form an index of a company's communication activities, covering aspects such as online advertising, sponsoring or product placement (Sarstedt and Mooi, 2019). PLS-SEM should always be the preferred approach in situations with formatively measured constructs, because a MIMIC approach in CB-SEM imposes constraints on the model that often contradict the theoretical assumptions (Sarstedt *et al.*, 2016b).

Statistical power

When using PLS-SEM, researchers benefit from the method's high degree of statistical power compared to CB-SEM (Reinartz *et al.*, 2009; Hair *et al.*, 2017b). This characteristic holds even when estimating common factor model data as assumed by CB-SEM (Sarstedt *et al.*, 2016b). Greater statistical power means that PLS-SEM is more likely to identify relationships as significant when they are indeed present in the population (Sarstedt and Mooi, 2019).

The PLS-SEM characteristic of higher statistical power is quite useful for exploratory research that examines less developed or still developing theory. Wold (1985, p. 590) describes the use of PLS-SEM as "a dialogue between the investigator and the computer. Tentative improvements of the model–such as the introduction of a new latent variable, an indicator, or an inner relation, or the omission of such an element–are tested for predictive relevance [. . .] and the various pilot studies are a speedy and low-cost matter." Of particular importance, however, is that PLS-SEM is not only appropriate for exploratory research but also for confirmatory research (Hair *et al.*, 2017a).

Goodness-of-fit

While CB-SEM strongly relies on the concept of model fit, this is much less the case with PLS-SEM (Hair *et al.*, 2019). Consequently, some researchers incorrectly conclude that PLS-SEM is not useful for theory testing and confirmation (Westland, 2015). A couple of methodologists have endorsed model fit measures for PLS-SEM (Henseler *et al.*, 2016a), but researchers should be very cautious when considering the applicability of these measures for PLS-SEM (Henseler and Sarstedt, 2013; Hair *et al.*, 2019). First, a comprehensive assessment of these measures has not been conducted so far. Therefore, any thresholds (guidelines) advocated in the literature should be considered as very tentative. Second, as the algorithm for obtaining PLS-SEM solutions is not based on minimizing the divergence between observed and estimated covariance matrices, the concept of Chi-square-based model fit measures and their extentions – as used in CB-SEM – are not applicable. Hence, even bootstrap-based model fit assessments on the grounds of, for example, some distance measure or the SRMR (Henseler *et al.*, 2016a; Henseler *et al.*, 2017), which quantify the divergence between the observed and estimated covariance matrices, should be considered with extreme caution. Third, scholars have questioned whether the concept of model fit, as applied in the context of CB-SEM research, is of value to PLS-SEM applications in general (Hair *et al.*, 2017a; Rigdon, 2012; Lohmöller, 1989).

PLS-SEM primarily focuses on the interplay between prediction and theory testing and results should be validated accordingly (Shmueli, 2010). In this context, scholars have recently proposed new evaluation procedures that are designed specifically for PLS-SEM's prediction-oriented nature (Shmueli *et al.*, 2016).

Evaluation of partial least squares-structural equation modeling results

The first step in evaluating PLS-SEM results involves examining the measurement models. The relevant criteria differ for reflective and formative constructs. If the measurement models meet all the required criteria, researchers then need to assess the structural model (Hair *et al.*, 2017a). As with most statistical methods, PLS-SEM has rules of thumb that serve as guidelines to evaluate model results (Chin, 2010; Götz *et al.*, 2010; Henseler *et al.*, 2009; Chin, 1998; Tenenhaus *et al.*, 2005; Roldán and Sánchez-Franco, 2012; Hair *et al.*, 2017a). Rules of thumb – by their very nature – are broad guidelines that suggest how to interpret the results, and they typically vary depending on the context. As an example, reliability for exploratory research should be a minimum of 0.60, while reliability for research that depends on established measures should be 0.70 or higher. The final step in interpreting PLS-SEM results, therefore, involves running one or more robustness checks to support the stability of results. The relevance of these robustness checks depends on the research context, such as the aim of the analysis and the availability of data.

Assessing reflective measurement models

The first step in reflective measurement model assessment involves examining the indicator loadings. Loadings above 0.708 are recommended, as they indicate that the construct explains more than 50 per cent of the indicator's variance, thus providing acceptable item reliability.

The second step is assessing internal consistency reliability, most often using Jöreskog's (1971) composite reliability. Higher values generally indicate higher levels of reliability. For example, reliability values between 0.60 and 0.70 are considered "acceptable in exploratory research," values between 0.70 and 0.90 range from "satisfactory to good." Values of 0.95 and higher are problematic, as they indicate that the items are redundant, thereby reducing construct validity (Diamantopoulos *et al.*, 2012; Drolet and Morrison, 2001). Reliability values of 0.95 and above also suggest the possibility of undesirable response patterns (e.g. straight lining), thereby triggering inflated correlations among the indicators' error terms. Cronbach's alpha is another measure of internal consistency reliability that assumes similar thresholds, but produces lower values than composite reliability. Specifically, Cronbach's alpha is a less precise measure of reliability, as the items are unweighted. In contrast, with composite reliability, the items are weighted based on the construct indicators' individual loadings and, hence, this reliability is higher than Cronbach's alpha. While Cronbach's alpha may be too conservative, the composite reliability may be too liberal, and the construct's true reliability is typically viewed as within these two extreme values. As an alternative, Dijkstra and Henseler (2015) proposed ρ_A as an approximately exact measure of construct reliability, which usually lies between Cronbach's alpha and the composite reliability. Hence, ρ_A may represent a good compromise if one assumes that the factor model is correct.

In addition, researchers can use bootstrap confidence intervals to test if the construct reliability is significantly higher than the recommended minimum threshold (e.g. the lower bound of the 95 per cent confidence interval of the construct reliability is higher than 0.70). Similarly, they can test if construct reliability is significantly lower than the recommended maximum threshold (e.g. the upper bound of the 95 per cent confidence interval of the construct reliability is lower than 0.95). To obtain the bootstrap confidence intervals, in line with Aguirre-Urreta and Rönkkö (2018), researchers should generally use the percentile method. However, when the reliability coefficient's bootstrap distribution is skewed, the BCa method should be preferred to obtain bootstrap confidence intervals.

The third step of the reflective measurement model assessment addresses the convergent validity of each construct measure. Convergent validity is the extent to which the construct converges to explain the variance of its items. The metric used for evaluating a construct's convergent validity is the average variance extracted (AVE) for all items on each construct. To calculate the AVE, one has to square the loading of each indicator on a construct and compute the mean value. An acceptable AVE is 0.50 or higher indicating that the construct explains at least 50 per cent of the variance of its items.

The fourth step is to assess discriminant validity, which is the extent to which a construct is empirically distinct from other constructs in the structural model. Fornell and Larcker (1981) proposed the traditional metric and suggested that each construct's AVE should be compared to the squared inter-construct correlation (as a measure of shared variance) of that same construct and all other reflectively measured constructs in the structural model. The shared variance for all model constructs should not be larger than their AVEs. Recent research indicates, however, that this metric is not suitable for discriminant validity assessment. For example, Henseler *et al.* (2015) show that the Fornell-Larcker criterion does not perform well, particularly when the indicator loadings on a construct differ only slightly (e.g. all the indicator loadings are between 0.65 and 0.85).

As a replacement, Henseler *et al.* (2015) proposed the heterotrait-monotrait (HTMT) ratio of the correlations (Voorhees *et al.*, 2016). The HTMT is defined as the mean value of the item correlations across constructs relative to the (geometric) mean of the average correlations for the items measuring the same construct. Discriminant validity problems are present when HTMT values are high. Henseler *et al.* (2015) propose a threshold value of 0.90 for structural models with constructs that are conceptually very similar, for instance cognitive satisfaction, affective satisfaction and loyalty. In such a setting, an HTMT value above 0.90 would suggest that discriminant validity is not present. But when constructs are conceptually more distinct, a lower, more conservative, threshold value is suggested, such as 0.85 (Henseler *et al.*, 2015). In addition to these guidelines, bootstrapping can be applied to test whether the HTMT value is significantly different from 1.00 (Henseler *et al.*, 2015) or a lower threshold value such as 0.85 or 0.90, which should be defined based on the study context (Franke and Sarstedt, 2019). More specifically, the researcher can examine if the upper bound of the 95 per cent confidence interval of HTMT is lower than 0.90 or 0.85.

Assessing formative measurement models

PLS-SEM is the preferred approach when formative constructs are included in the structural model (Hair *et al.*, 2019). Formative measurement models are evaluated based on the following: convergent validity, indicator collinearity, statistical significance, and relevance of the indicator weights (Hair *et al.*, 2017a).

For formatively measured constructs, convergent validity is assessed by the correlation of the construct with an alternative measure of the same concept. Originally proposed by Chin (1998), the procedure is referred to as redundancy analysis. To execute this procedure for determining convergent validity, researchers must plan already in the research design stage to include alternative reflectively measured indicators of the same concept in their questionnaire. Cheah *et al.* (2018) show that a single-item, which captures the essence of the construct under consideration, is generally sufficient as an alternative measure – despite limitations with regard to criterion validity (Sarstedt *et al.*, 2016a). When the model is based on secondary data, a variable measuring a similar concept would be used (Houston, 2004). Hair *et al.* (2017a) suggest that the correlation of the formatively measured construct with the single-item construct, measuring the same concept, should be 0.70 or higher.

The variance inflation factor (VIF) is often used to evaluate collinearity of the formative indicators. VIF values of 5 or above indicate critical collinearity issues among the indicators of formatively measured constructs. However, collinearity issues can also occur at lower VIF values of 3 (Mason and Perreault, 1991; Becker *et al.*, 2015). Ideally, the VIF values should be close to 3 and lower.

In the third and final step, researchers need to assess the indicator weights' statistical significance and relevance (i.e. size). PLS-SEM is a nonparametric method and therefore, bootstrapping is used to determine statistical significance (Chin, 1998). Hair *et al.* (2017a) suggest using BCa bootstrap confidence intervals for significance testing in case the bootstrap distribution of the indicator weights is skewed. Otherwise, researchers should use the percentile method to construct bootstrap-based confidence intervals (Aguirre-Urreta and Rönkkö, 2018). If the confidence interval of an indicator weight includes zero, this indicates that the weight is not statistically significant and the indicator should be considered for removal from the measurement model. However, if an indicator weight is not significant, it is not necessarily interpreted as evidence of poor measurement model quality. Instead, the indicator's absolute contribution to the construct is considered (Cenfetelli and Bassellier, 2009), as defined by its outer loading (i.e. the bivariate correlation between the indicator and its construct). According to Hair *et al.* (2017a), indicators with a nonsignificant weight should definitely be eliminated if the loading is also not significant. A low but significant loading of 0.50 and below suggests that one should consider deleting the indicator, unless there is strong support for its inclusion on the grounds of measurement theory.

When deciding whether to delete formative indicators based on statistical outcomes, researchers need to be cautious for the following reasons. First, formative indicator weights are a function of the number of indicators used to measure a construct. The greater the number of indicators, the lower their average weight. Formative measurement models are, therefore, inherently limited in the number of indicator weights that can be statistically significant (Cenfetelli and Bassellier, 2009). Second, indicators should seldom be removed from formative measurement models, as formative measurement theory requires the indicators to fully capture the entire domain of a construct, as defined by the researcher in the conceptualization stage. In contrast to reflective measurement models, formative indicators are not interchangeable and removing even a single indicator can therefore, reduce the measurement model's content validity (Diamantopoulos and Winklhofer, 2001).

After assessing the statistical significance of the indicator weights, researchers need to examine each indicator's relevance. The indicator weights are standardized to values between −1 and +1, but, in rare cases can also take values lower or higher than this, which indicates an abnormal result (e.g. due to collinearity issues and/or small sample sizes). A weight close to 0 indicates a weak relationship, whereas weights close to +1 (or −1) indicate strong positive (or negative) relationships.

Assessing structural models

When the measurement model assessment is satisfactory, the next step in evaluating PLS-SEM results is assessing the structural model. Standard assessment criteria, which should be considered, include the coefficient of determination (R^2), the blindfolding-based cross-validated redundancy measure Q^2, and the statistical significance and relevance of the path coefficients. In addition, researchers should assess their model's out-of-sample predictive power by using the PLSpredict procedure (Shmueli *et al.*, 2016).

Structural model coefficients for the relationships between the constructs are derived from estimating a series of regression equations. Before assessing the structural relationships, collinearity must be examined to make sure it does not bias the regression results. This process is similar to assessing formative measurement models, but the latent variable scores of the predictor constructs in a partial regression are used to calculate the VIF values. VIF values above 5 are indicative of probable collinearity issues among the predictor constructs, but collinearity problems can also occur at lower VIF values of 3-5 (Mason and Perreault, 1991; Becker *et al.*, 2015). Ideally, the VIF values should be close to 3 and lower. If collinearity is a problem, a frequently used option is to create higher-order models that can be supported by theory (Hair *et al.*, 2017a).

If collinearity is not an issue, the next step is examining the R^2 value of the endogenous construct(s). The R^2 measures the variance, which is explained in each of the endogenous constructs and is therefore a measure of the model's explanatory power (Shmueli and Koppius, 2011). The R^2 is also referred to as in-sample predictive power (Rigdon, 2012). The R^2 ranges from 0 to 1, with higher values indicating a greater explanatory power. As a guideline, R^2 values of 0.75, 0.50 and 0.25 can be considered substantial, moderate and weak (Henseler *et al.*, 2009; Hair *et al.*, 2011). Acceptable R^2 values are based on the context and in some disciplines an R^2 value as low as 0.10 is considered satisfactory, for example, when predicting stock returns (Raithel *et al.*, 2012). More importantly, the R^2 is a function of the number of predictor constructs – the greater the number of predictor constructs, the higher the R^2. Therefore, the R^2 should always be interpreted in relation to the context of the study, based on the R^2 values from related studies and models of similar complexity. R^2 values can also be too high when the model overfits the data. That is, the partial regression model is too complex, which results in fitting the random noise inherent in the sample rather than reflecting the overall population. The same model would likely not fit on another sample drawn from the same population (Sharma *et al.*, 2019a). When measuring a concept that is inherently predictable, such as physical processes, R^2 values of 0.90 might be plausible. Similar R^2 value levels in a model that predicts human attitudes, perceptions and intentions likely indicate an overfit.

Researchers can also assess how the removal of a certain predictor construct affects an endogenous construct's R^2 value. This metric is the f^2 effect size and is somewhat redundant to the size of the path coefficients. More precisely, the rank order of the predictor constructs' relevance in explaining a dependent construct in the structural model is often the same when comparing the size of the path coefficients and the f^2 effect sizes. In such situations, the f^2 effect size should only be reported if requested by editors or reviewers. If the rank order of the constructs' relevance, when explaining a dependent construct in the structural model, differs when comparing the size of the path coefficients and the f^2 effect sizes, the researcher may report the f^2 effect size to explain the presence of, for example, partial or full mediation (Nitzl *et al.*, 2016). As a rule of thumb, values higher than 0.02, 0.15 and 0.35 depict small, medium and large f^2 effect sizes (Cohen, 1988).

Another means to assess the PLS path model's predictive accuracy is by calculating the Q^2 value (Geisser, 1974; Stone, 1974). This metric is based on the blindfolding procedure that removes single points in the data matrix, imputes the removed points with the mean and estimates the model parameters (Rigdon, 2014b; Sarstedt *et al.*, 2014). As such, the Q^2 is not a measure of out-of-sample prediction, but rather combines aspects of out-of-sample prediction and in-sample explanatory power (Shmueli *et al.*, 2016; Sarstedt *et al.*, 2017a). Using these estimates as input, the blindfolding procedure predicts the data points that were removed for all variables. Small differences between the predicted and the original values translate into a higher Q^2 value, thereby indicating a higher predictive accuracy. As a guideline, Q^2 values should be larger than zero for a specific endogenous construct to indicate predictive accuracy of the structural model for that construct. As a rule of thumb, Q^2 values higher than 0, 0.25 and 0.50 depict small, medium and large predictive relevance of the PLS-path model. Similar to the f^2 effect sizes, it is possible to compute and interpret the q^2 effect sizes.

Many researchers interpret the R^2 statistic as a measure of their model's predictive power. This interpretation is not entirely correct, however, as the R^2 only indicates the model's in-sample explanatory power – it says nothing about the model's out-of-sample predictive power (Shmueli, 2010; Shmueli and Koppius, 2011; Dolce *et al.*, 2017). Addressing this concern, Shmueli *et al.* (2016) proposed a set of procedures for out-of-sample prediction that involves estimating the model on an analysis (i.e. training) sample and evaluating its predictive performance on data other than the analysis sample, referred to as a holdout sample. Their PLSpredict procedure generates holdout sample-based predictions in PLS-SEM and is an option in PLS-SEM software, such as SmartPLS (Ringle *et al.*, 2015) and open source environments such as R (https://github.com/ISS-Analytics/pls-predict), so that researchers can easily apply the procedure.

PLSpredict executes k-fold cross-validation. A fold is a subgroup of the total sample and k is the number of subgroups. That is, the total data set is randomly split into k equally sized subsets of data. For example, a cross-validation based on $k = 5$ folds splits the sample into five equally sized data subsets (i.e. groups of data). PLSpredict then combines $k - 1$ subsets into a single analysis sample that is used to predict the remaining fifth data subset. The fifth data subset is the holdout sample for the first cross-validation run. This cross-validation process is then repeated k times (in this example, five times), with each of the five subsets used once as the holdout sample. Thus, each case in every holdout sample has a predicted value estimated with a sample in which that case was not used to estimate the model parameters. Shmueli *et al.* (2019) recommend setting $k = 10$, but researchers need to make sure the analysis sample for each subset (fold) meets minimum sample size guidelines. Also, other criteria to assess out-of-sample prediction without using a holdout sample are available, such as the Bayesian information criterion (BIC) and Geweke and Meese (GM) criterion (discussed later in this paper).

The generation of the k subgroups is a random process and can sometimes result in extreme partitions that potentially lead to abnormal solutions. To avoid such abnormal solutions, researchers should run PLSpredict multiple times. Shmueli *et al.* (2019) recommend to generally run the procedure ten times. However, when the objective is to duplicate how the PLS model will eventually be used to predict a new observation by using a single model (estimated from the entire data set), PLSpredict should be run only once (i.e. without repetitions).

For the PLSpredict based assessment of a model's predictive power, researchers can draw on several prediction statistics that quantify the amount of prediction error. For example, the mean absolute error (MAE) measures the average magnitude of the errors in a set of predictions without considering their direction (over or under). The MAE is thus the average absolute differences between the predictions and the actual observations, with all the individual differences having equal weight. Another popular prediction metric is the root mean squared error (RMSE), which is defined as the square root of the average of the squared differences between the predictions and the actual observations. As the RMSE squares the errors before averaging, the statistic assigns a greater weight to larger errors, which makes it particularly useful when large errors are undesirable – as is typically the case in business research applications.

When interpreting PLSpredict results, the focus should be on the model's key endogenous construct, as opposed to examining the prediction errors for all endogenous constructs' indicators. When the key target construct has been selected, the Q^2_{predict} statistic should be evaluated first to verify if the predictions outperform the most naïve benchmark, defined as the indicator means from the analysis sample (Shmueli *et al.*, 2019). Then, researchers need to examine the prediction statistics. In most instances, researchers should use the RMSE. If the prediction error distribution is highly non-symmetric, the MAE is the more appropriate prediction statistic (Shmueli *et al.*, 2019). The prediction statistics depend on the indicators' measurement scales and their raw values do not carry much meaning. Therefore, researchers need to compare the RMSE (or MAE) values with a naïve benchmark. The recommended naïve benchmark (produced by the PLSpredict method) uses a linear regression model (LM) to generate predictions for the manifest variables, by running a linear regression of each of the dependent construct's indicators on the indicators of the exogenous latent variables in the PLS path model (Danks and Ray, 2018). When comparing the RMSE (or MAE) values with the LM values, the following guidelines apply (Shmueli *et al.*, 2019):

- If the PLS-SEM analysis, compared to the naïve LM benchmark, yields higher prediction errors in terms of RMSE (or MAE) for *all* indicators, this indicates that the model lacks predictive power.
- If the majority of the dependent construct indicators in the PLS-SEM analysis produce higher prediction errors compared to the naïve LM benchmark, this indicates that the model has a low predictive power.
- If the minority (or the same number) of indicators in the PLS-SEM analysis yields higher prediction errors compared to the naïve LM benchmark, this indicates a medium predictive power.
- If none of the indicators in the PLS-SEM analysis has higher RMSE (or MAE) values compared to the naïve LM benchmark, the model has high predictive power.

Having substantiated the model's explanatory power and predictive power, the final step is to assess the statistical significance and relevance of the path coefficients. The interpretation of the path coefficients parallels that of the formative indicator weights. That is, researchers need to run bootstrapping to assess the path coefficients' significance and evaluate their values, which typically fall in the range of -1 and $+1$. Also, they can interpret a construct's indirect effect on a certain target construct via one or more intervening constructs. This effect type is particularly relevant in the assessment of mediating effects (Nitzl, 2016).

Similarly, researchers can interpret a construct's total effect, defined as the sum of the direct and all indirect effects. A model's total effects also serve as input for the importance-performance map analysis (IPMA) and extend the standard PLS-SEM results reporting of path coefficient estimates by adding a dimension to the analysis that considers the average values of the latent variable scores. More precisely, the IPMA compares the structural model's total effects on a specific target construct with the average latent variable scores of this construct's predecessors (Ringle and Sarstedt, 2016).

Finally, researchers may be interested in comparing different model configurations resulting from different theories or research contexts. Sharma *et al.* (2019b, 2019a) recently compared the efficacy of various metrics for model comparison tasks and found that Schwarz's (1978) BIC and Geweke and Meese's (1981) GM achieve a sound trade-off between model fit and predictive power in the estimation of PLS path models. Their research facilitates assessing out-of-sample prediction without using a holdout sample, and is particularly useful with PLS-SEM applications based on a sample that is too small to divide it into useful analysis and holdout samples. Specifically, researchers should estimate each model separately and select the model that minimizes the value in BIC or GM for a certain target construct. For example, a model that produces a BIC value of −270 should be preferred over a model that produces a BIC value of −150. Table I summarizes the metrics that need to be applied when interpreting and reporting PLS-SEM results.

Robustness checks

Recent research has proposed complementary methods for assessing the robustness of PLS-SEM results (Hair *et al.*, 2018; Latan, 2018). These methods address either the measurement model or the structural model (Table I).

In terms of measurement models, Gudergan *et al.* (2008) have proposed the confirmatory tetrad analysis (CTA-PLS), which enables empirically substantiating the specification of measurement models (i.e. reflective versus formative). The CTA-PLS relies on the concept of tetrads that describe the difference of the product of one pair of covariances and the product of another pair of covariances (Bollen and Ting, 2000). In a reflective measurement model, these tetrads should vanish (i.e. they become zero) as the indicators are assumed to stem from the same domain. If one of a construct's tetrads is significantly different from 0, one rejects the null hypothesis and assumes a formative instead of a reflective measurement model specification. It should be noted, however, that CTA-PLS is an empirical test of measurement models and the primary method to determine reflective or formative model specification is theoretical reasoning (Hair *et al.*, 2017a).

In terms of the structural model, Sarstedt *et al.* (2019) suggest that researchers should consider nonlinear effects, endogeneity and unobserved heterogeneity. First, to test whether relationships are nonlinear, researchers can run Ramsey's (1969) regression equation specification error test on the latent variable scores in the path model's partial regressions. A significant test statistic in any of the partial regressions indicates a potential nonlinear effect. In addition, researchers can establish an interaction term to map a nonlinear effect in the model and test its statistical significance using bootstrapping (Svensson *et al.*, 2018).

Second, when the research perspective is primarily explanatory in a PLS-SEM analysis, researchers should test for endogeneity. Endogeneity typically occurs when researchers have omitted a construct that correlates with one or more predictor constructs and the dependent construct in a partial regression of the PLS path model. To assess and treat endogeneity, researchers should follow Hult *et al.*'s (2018) systematic procedure, starting with the application of Park and Gupta's (2012) Gaussian copula approach. If the approach indicates an endogeneity issue, researchers should implement instrumental variables that are highly correlated with the independent constructs, but are uncorrelated with the dependent construct's error term to explain the sources of endogeneity (Bascle, 2008). Importantly, however, endogeneity assessment is only relevant when the researcher's focus is on explanation and rather not when following causal-predictive goals.

Third, unobserved heterogeneity occurs when subgroups of data exist that produce substantially different model estimates. If this is the case, estimating the model based on the entire data set is very likely to produce misleading results (Becker *et al.*, 2013). Hence, any

Reflective measurement models	
Reflective indicator loadings	$\geq$0.708
Internal consistency reliability	Cronbach's alpha is the lower bound, the composite reliability is the upper bound for internal consistency reliability. ρ_A usually lies between these bounds and may serve as a good representation of a construct's internal consistency reliability, assuming that the factor model is correct Minimum 0.70 (or 0.60 in exploratory research) Maximum of 0.95 to avoid indicator redundancy, which would compromise content validity Recommended 0.70-0.90 Test if the internal consistency reliability is significantly higher (lower) than the recommended minimum (maximum) thresholds. Use the percentile method to construct the bootstrap-based confidence interval; in case of a skewed bootstrap distribution, use the BCa method
Convergent validity	AVE $\geq$ 0.50
Discriminant validity	For conceptually similar constructs: HTMT $<$ 0.90 For conceptually different constructs: HTMT $<$ 0.85 Test if the HTMT is significantly lower than the threshold value
Formative measurement models	
Convergent validity (redundancy analysis)	$\geq$0.70 correlation
Collinearity (VIF)	Probable (i.e. critical) collinearity issues when VIF $\geq$ 5 Possible collinearity issues when VIF $\geq$ 3-5 Ideally show that VIF $<$ 3
Statistical significance of weights	p-value $<$ 0.05 or the 95% confidence interval (based on the percentile method or, in case of a skewed bootstrap distribution, the BCa method) does not include zero
Relevance of indicators with a significant weight	Larger significant weights are more relevant (contribute more)
Relevance of indicators with a non-significant weight	Loadings of $\geq$0.50 that are statistically significant are considered relevant
Structural model	
Collinearity (VIF)	Probable (i.e. critical) collinearity issues when VIF $\geq$ 5 Possible collinearity issues when VIF $\geq$ 3-5 Ideally show that VIF $<$ 3
R^2 value	R^2 values of 0.75, 0.50 and 0.25 are considered substantial, moderate and weak. R^2 values of 0.90 and higher are typical indicative of overfit
Q^2 value	Values larger than zero are meaningful Values higher than 0, 0.25 and 0.50 depict small, medium and large predictive accuracy of the PLS path model
PLSpredict	Set k = 10, assuming each subgroup meets the minimum required sample size Use ten repetitions, assuming the sample size is large enough $Q^2_{predict}$ values $>$ 0 indicate that the model outperforms the most naïve benchmark (i.e. the indicator means from the analysis sample) Compare the MAE (or the RMSE) value with the LM value of each

(*continued*)

Table I.
Guidelines when using PLS-SEM

	indicator. Check if the PLS-SEM analysis (compared to the LM) yields higher prediction errors in terms of RMSE (or MAE) for all (no predictive power), the majority (low predictive power), the minority or the same number (medium predictive power) or none of the indicators (high predictive power)
Model comparisons	Select the model that minimizes the value in BIC or GM compared to the other models in the set
Robustness checks	
Measurement models	CTA-PLS
Structural model	Nonlinear effects
	Endogeneity
	Unobserved heterogeneity

Table I.

PLS-SEM analysis should include a routine check for unobserved heterogeneity to ascertain whether or not the analysis of the entire data set is reasonable or not. Sarstedt *et al.* (2017b) proposed a systematic procedure for identifying and treating unobserved heterogeneity. Using information criteria derived from a finite mixture PLS (Hahn *et al.*, 2002; Sarstedt *et al.*, 2011), researchers can identify the number of segments to be extracted from the data (if any) (Hair *et al.*, 2016; Matthews *et al.*, 2016). If heterogeneity is present at a critical level, the next step involves running the PLSprediction-oriented segmentation procedure (Becker *et al.*, 2013) to disclose the data's segment structure. Finally, researchers should attempt to identify suitable explanatory variables that characterize the uncovered segments (e.g. by using contingency table or exhaustive CHAID analyses; Ringle *et al.*, 2010). If suitable explanatory variables are available, a moderator (Henseler and Fassott, 2010; Becker *et al.*, 2018) or multigroup analysis (Chin and Dibbern, 2010; Matthews, 2017), in combination with a measurement invariance assessment (Henseler *et al.*, 2016b), offers further particularized findings, conclusions and implications.

Concluding observations

PLS-SEM is increasingly being applied to estimate structural equation models (Hair *et al.*, 2014). Scholars need a comprehensive, yet concise, overview of the considerations and metrics needed to ensure their analysis and reporting of PLS-SEM results is complete – before submitting their article for review. Prior research has provided such reporting guidelines (Hair *et al.*, 2011; Hair *et al.*, 2013; Hair *et al.*, 2012b; Chin, 2010; Tenenhaus *et al.*, 2005; Henseler *et al.*, 2009), which, in light of more recent research and methodological developments in the PLS-SEM domain, need to be continuously extended and updated. We hope this paper achieves this goal.

For researchers who have not used PLS-SEM in the past, this article is a good point of orientation on when preparing and finalizing their manuscripts. Moreover, for researchers experienced in applying PLS-SEM, this is a good overview and reminder of how to prepare PLS-SEM manuscripts. This knowledge is also important for reviewers and journal editors to ensure the rigor of published PLS-SEM studies. We provide an overview of several recently proposed improvements (PLSpredict and model comparison metrics), as well as complementary methods for robustness checks (e.g. endogeneity assessment and latent class procedures), which we recommend should be applied – if appropriate – when using PLS-SEM. Finally, while a few researchers have published articles that are negative about the use of PLS-SEM, more recently several prominent researchers have acknowledged the value of PLS as an SEM technique (Petter, 2018). We believe that social science scholars would be remiss if they did not apply all statistical methods at their disposal to explore and better understand the phenomena they are researching.

References

Aguirre-Urreta, M.I. and Rönkkö, M. (2018), "Statistical inference with PLSc using bootstrap confidence intervals", *MIS Quarterly*, Vol. 42 No. 3, pp. 1001-1020.

Akter, S., Fosso Wamba, S. and Dewan, S. (2017), "Why PLS-SEM is suitable for complex modelling? An empirical illustration in big data analytics quality", *Production Planning and Control*, Vol. 28 Nos 11/12, pp. 1011-1021.

Ali, F., Rasoolimanesh, S.M. and Cobanoglu, C. (2018a), *Applying Partial Least Squares in Tourism and Hospitality Research*, Emerald, Bingley.

Ali, F., Rasoolimanesh, S.M., Sarstedt, M., Ringle, C.M. and Ryu, K. (2018b), "An assessment of the use of partial least squares structural equation modeling (PLS-SEM) in hospitality research", *International Journal of Contemporary Hospitality Management*, Vol. 30 No. 1, pp. 514-538.

Avkiran, N.K. and Ringle, C.M. (2018), *Partial Least Squares Structural Equation Modeling: Recent Advances in Banking and Finance*, Springer International Publishing, Cham.

Bascle, G. (2008), "Controlling for endogeneity with instrumental variables in strategic management research", *Strategic Organization*, Vol. 6 No. 3, pp. 285-327.

Becker, J.-M., Ringle, C.M. and Sarstedt, M. (2018), "Estimating moderating effects in PLS-SEM and PLSc-SEM: interaction term generation*data treatment", *Journal of Applied Structural Equation Modeling*, Vol. 2 No. 2, pp. 1-21.

Becker, J.-M., Rai, A., Ringle, C.M. and Völckner, F. (2013), "Discovering unobserved heterogeneity in structural equation models to avert validity threats", *MIS Quarterly*, Vol. 37 No. 3, pp. 665-694.

Becker, J.-M., Ringle, C.M., Sarstedt, M. and Völckner, F. (2015), "How collinearity affects mixture regression results", *Marketing Letters*, Vol. 26 No. 4, pp. 643-659.

Bollen, K.A. and Ting, K.-F. (2000), "A tetrad test for causal indicators", *Psychological Methods*, Vol. 5 No. 1, pp. 3-22.

Boomsma, A. and Hoogland, J.J., (2001), "The robustness of LISREL modeling revisited", in Cudeck, R., du Toit, S. and Sörbom, D. (Eds) *Structural Equation Modeling: Present and Future*, Scientific Software International, Chicago. 139-168.

Cenfetelli, R.T. and Bassellier, G. (2009), "Interpretation of formative measurement in information systems research", *MIS Quarterly*, Vol. 33 No. 4, pp. 689-708.

Cheah, J.-H., Sarstedt, M., Ringle, C.M., Ramayah, T. and Ting, H. (2018), "Convergent validity assessment of formatively measured constructs in PLS-SEM: on using single-item versus multi-item measures in redundancy analyses", *International Journal of Contemporary Hospitality Management*, Vol. 30 No. 11, pp. 3192-3210.

Chin, W.W. (1998), "The partial least squares approach to structural equation modeling", in Marcoulides, G.A. (Ed.), *Modern Methods for Business Research*, Mahwah, Erlbaum, pp. 295-358.

Chin, W.W. (2003), *PLS-Graph 3.0*, Soft Modeling Inc, Houston.

Chin, W.W. (2010), "How to write up and report PLS analyses", in Esposito Vinzi, V., Chin, W.W., Henseler, J., *et al.* (Eds), *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series*), Springer, Heidelberg, Dordrecht, London, New York, NY, Vol. II, pp. 655-690.

Chin, W.W. and Dibbern, J. (2010), "A permutation based procedure for multi-group PLS analysis: results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA", in Esposito Vinzi, V., Chin, W.W., Henseler J. and Wang, H. (Eds), *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series)*, Springer, Heidelberg, Dordrecht, London, New York, NY, Vol. II, pp. 171-193.

Chou, C.-P., Bentler, P.M. and Satorra, A. (1991), "Scaled test statistics and robust standard errors for Non-Normal data in covariance structure analysis: a monte carlo study", *British Journal of Mathematical and Statistical Psychology*, Vol. 44 No. 2, pp. 347-357.

Cochran, W.G. (1977), *Sampling Techniques*, Wiley, New York, NY.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences: Lawrence Erlbaum Associates*.

Danks, N. and Ray, S. (2018), "Predictions from partial least squares models", in Ali, F., Rasoolimanesh, S.M. and Cobanoglu, C. (Eds), *Applying Partial Least Squares in Tourism and Hospitality Research*, Emerald, Bingley, pp. 35-52.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P. and Kaiser, S. (2012), "Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective", *Journal of the Academy of Marketing Science*, Vol. 40 No. 3, pp. 434-449.

Diamantopoulos, A. and Winklhofer, H.M. (2001), "Index construction with formative indicators: an alternative to scale development", *Journal of Marketing Research*, Vol. 38 No. 2, pp. 269-277.

Dijkstra, T.K. and Henseler, J. (2015), "Consistent partial least squares path modeling", *MIS Quarterly*, Vol. 39 No. 2, pp. 297-316.

do Valle, P.O. and Assaker, G. (2016), "Using partial least squares structural equation modeling in tourism research: a review of past research and recommendations for future applications", *Journal of Travel Research*, Vol. 55 No. 6, pp. 695-708.

Dolce, P., Esposito Vinzi, V. and Lauro, C. (2017), "Predictive path modeling through PLS and other component-based approaches: methodological issues and performance evaluation", in Latan, H. and Noonan, R. (Eds), *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications*, Springer International Publishing, Cham, pp. 153-172.

Drolet, A.L. and Morrison, D.G. (2001), "Do we really need multiple-item measures in service research?", *Journal of Service Research*, Vol. 3 No. 3, pp. 196-204.

Efron, B. (1987), "Better bootstrap confidence intervals", *Journal of the American Statistical Association*, Vol. 82 No. 397, pp. 171-185.

Fornell, C.G. and Bookstein, F.L. (1982), "Two structural equation models: LISREL and PLS applied to consumer exit-voice theory", *Journal of Marketing Research*, Vol. 19 No. 4, pp. 440-452.

Fornell, C.G. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Franke, G.R. and Sarstedt, M. (2019), "Heuristics versus statistics in discriminant validity testing: a comparison of four procedures", *Internet Research*, Forthcoming.

Garson, G.D. (2016), *Partial Least Squares Regression and Structural Equation Models*, Statistical Associates, Asheboro.

Geisser, S. (1974), "A predictive approach to the random effects model", *Biometrika*, Vol. 61 No. 1, pp. 101-107.

Geweke, J. and Meese, R. (1981), "Estimating regression models of finite but unknown order", *International Economic Review*, Vol. 22 No. 1, pp. 55-70.

Goodhue, D.L., Lewis, W. and Thompson, R. (2012), "Does PLS have advantages for small sample size or non-normal data?", *MIS Quarterly*, Vol. 36 No. 3, pp. 981-1001.

Götz, O., Liehr-Gobbers, K. and Krafft, M. (2010), "Evaluation of structural equation models using the partial least squares (PLS) Approach", in Esposito Vinzi, V., Chin, W.W., Henseler, J., *et al.* (Eds), *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series)*, Springer, Heidelberg, Dordrecht, London, New York, NY, pp. 691-711.

Gudergan, S.P., Ringle, C.M., Wende, S. and Will, A. (2008), "Confirmatory tetrad analysis in PLS path modeling", *Journal of Business Research*, Vol. 61 No. 12, pp. 1238-1249.

Hahn, C., Johnson, M.D., Herrmann, A. and Huber, F. (2002), "Capturing customer heterogeneity using a finite mixture PLS approach", *Schmalenbach Business Review*, Vol. 54 No. 3, pp. 243-269.

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2017a), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage, Thousand Oaks, CA.

Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M. and Thiele, K.O. (2017b), "Mirror, Mirror on the wall: a comparative evaluation of composite-based structural equation modeling methods", *Journal of the Academy of Marketing Science*, Vol. 45 No. 5, pp. 616-632.

Hair, J.F., Ringle, C.M. and Sarstedt, M. (2011), "PLS-SEM: indeed a silver bullet", *Journal of Marketing Theory and Practice*, Vol. 19 No. 2, pp. 139-151.

Hair, J.F., Ringle, C.M. and Sarstedt, M. (2013), "Partial least squares structural equation modeling: rigorous applications, better results and higher acceptance", *Long Range Planning*, Vol. 46 Nos 1/2, pp. 1-12.

Hair, J.F., Sarstedt, M., Hopkins, L. and Kuppelwieser, V.G. (2014), "Partial least squares structural equation modeling (PLS-SEM): an emerging tool in business research", *European Business Review*, Vol. 26 No. 2, pp. 106-121.

Hair, J.F., Sarstedt, M., Matthews, L. and Ringle, C.M. (2016), "Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I – method", *European Business Review*, Vol. 28 No. 1, pp. 63-76.

Hair, J.F., Sarstedt, M., Pieper, T.M. and Ringle, C.M. (2012a), "The use of partial least squares structural equation modeling in strategic management research: a review of past practices and recommendations for future applications", *Long Range Planning*, Vol. 45 Nos 5/6, pp. 320-340.

Hair, J.F., Sarstedt, M. and Ringle, C.M. (2019), "Rethinking some of the rethinking of partial least squares", *European Journal of Marketing*, Forthcoming.

Hair, J.F., Sarstedt, M., Ringle, C.M. and Gudergan, S.P. (2018), *Advanced Issues in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage, Thousand Oaks, CA.

Hair, J.F., Sarstedt, M., Ringle, C.M., *et al.* (2012b), "An assessment of the use of partial least squares structural equation modeling in marketing research", *Journal of the Academy of Marketing Science*, Vol. 40 No. 3, pp. 414-433.

Henseler, J., Dijkstra, T.K., Sarstedt, M., Ringle, C.M., Diamantopoulos, A., Straub, D.W., Ketchen, D.J., Hair, J.F., Hult, G.T.M. and Calantone, R.J. (2014), "Common beliefs and reality about partial least squares: comments on Rönkkö and Evermann (2013)", *Organizational Research Methods*, Vol. 17 No. 2, pp. 182-209.

Henseler, J. and Fassott, G. (2010), "Testing moderating effects in PLS path models: an illustration of available procedures", in Esposito Vinzi, V., Chin, WW, Henseler, J., *et al.* (Eds), *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series)*, Springer, Heidelberg, Dordrecht, London, New York, NY, Vol. II, pp. 713-735.

Henseler, J., Hubona, G.S. and Ray, P.A. (2016a), "Using PLS path modeling in new technology research: Updated guidelines", *Industrial Management and Data Systems*, Vol. 116 No. 1, pp. 1-19.

Henseler, J., Hubona, G.S. and Ray, P.A. (2017), "Partial least squares path modeling: updated guidelines", in Latan, H. and Noonan, R. (Eds), *Partial Least Squares Structural Equation Modeling: Basic Concepts, Methodological Issues and Applications*, Springer, Heidelberg, pp. 19-39.

Henseler, J., Ringle, C.M. and Sarstedt, M. (2015), "A new criterion for assessing discriminant validity in variance-based structural equation modeling", *Journal of the Academy of Marketing Science*, Vol. 43 No. 1, pp. 115-135.

Henseler, J., Ringle, C.M. and Sarstedt, M. (2016b), "Testing measurement invariance of composites using partial least squares", *International Marketing Review*, Vol. 33 No. 3, pp. 405-431.

Henseler, J., Ringle, C.M. and Sinkovics, R.R. (2009), "The use of partial least squares path modeling in international marketing", in Sinkovics, R.R. and Ghauri, P.N. (Eds) *Advances in International Marketing*, Emerald, Bingley, pp. 277-320.

Henseler, J. and Sarstedt, M. (2013), "Goodness-of-fit indices for partial least squares path modeling", *Computational Statistics*, Vol. 28 No. 2, pp. 565-580.

Houston, M.B. (2004), "Assessing the validity of secondary data proxies for marketing constructs", *Journal of Business Research*, Vol. 57 No. 2, pp. 154-161.

Hult, G.T.M., Hair, J.F., Proksch, D., Sarstedt, M., Pinkwart, A. and Ringle, C.M. (2018), "Addressing endogeneity in international marketing applications of partial least squares structural equation modeling", *Journal of International Marketing*, Vol. 26 No. 3, pp. 1-21.

Ittner, C.D., Larcker, D.F. and Rajan, M.V. (1997), "The choice of performance measures in annual bonus contracts", *Accounting Review*, Vol. 72 No. 2, pp. 231-255.

Jöreskog, K.G. (1971), "Simultaneous factor analysis in several populations", *Psychometrika*, Vol. 36 No. 4, pp. 409-426.

Jöreskog, K.G. (1973), "A general method for estimating a linear structural equation system", In: Goldberger, A.S. and Duncan, O.D. (Eds), *Structural Equation Models in the Social Sciences*, Seminar Press, New York, NY, pp. 255-284.

Kaufmann, L. and Gaeckler, J. (2015), "A structured review of partial least squares in supply chain management research", *Journal of Purchasing and Supply Management*, Vol. 21 No. 4, pp. 259-272.

Khan, G.F., Sarstedt, M., Shiau, W.-L., Hair, J.F., Ringle, C.M. and Fritze, M. (2019), "Methodological research on partial least squares structural equation modeling (PLS-SEM): an analysis based on social network approaches", *Internet Research*, Forthcoming.

Kock, N. and Hadaya, P. (2018), "Minimum sample size estimation in PLS-SEM: the inverse square root and gamma-exponential methods", *Information Systems Journal*, Vol. 28 No. 1, pp. 227-261.

Latan, H. (2018), "PLS path modeling in hospitality and tourism research: the golden age and days of future Past", in Ali, F., Rasoolimanesh, S.M. and Cobanoglu, C. (Eds), *Applying Partial Least Squares in Tourism and Hospitality Research*, Emerald, Bingley, pp. 53-84.

Lohmöller, J.-B. (1989), *Latent Variable Path Modeling with Partial Least Squares*, Physica, Heidelberg.

Marcoulides, G.A. and Chin, W.W. (2013), "You write, but others read: common methodological misunderstandings in PLS and related methods", in Abdi, H., Chin, W.W., Esposito Vinzi, V., *et al.* (Eds), *New Perspectives in Partial Least Squares and Related Methods*, Springer, New York, NY, pp. 31-64.

Marcoulides, G.A., Chin, W.W. and Saunders, C. (2009), "Foreword: a critical look at partial least squares modeling", *MIS Quarterly*, Vol. 33 No. 1, pp. 171-175.

Marcoulides, G.A., Chin, W.W. and Saunders, C. (2012), "When imprecise statistical statements become problematic: a response to Goodhue, Lewis, and Thompson", *MIS Quarterly*, Vol. 36 No. 3, pp. 717-728.

Marcoulides, G.A. and Saunders, C. (2006), "PLS: a silver bullet?", *MIS Quarterly*, Vol. 30 No. 2, pp. III-IIX.

Mason, C.H. and Perreault, W.D. (1991), "Collinearity, power, and interpretation of multiple regression analysis", *Journal of Marketing Research*, Vol. 28 No. 3, pp. 268-280.

Mateos-Aparicio, G. (2011), "Partial least squares (PLS) methods: origins, evolution, and application to social sciences", *Communications in Statistics – Theory and Methods*, Vol. 40 No. 13, pp. 2305-2317.

Matthews, L. (2017), "Applying multi-group analysis in PLS-SEM: a step-by-step process", in Latan, H. and Noonan, R. (Eds), *Partial Least Squares Structural Equation Modeling: Basic Concepts, Methodological Issues and Applications*, Springer, Heidelberg, pp. 219-243.

Matthews, L., Sarstedt, M., Hair, J.F. and Ringle, C.M. (2016), "Identifying and treating unobserved heterogeneity with FIMIX-PLS: part II – a case study", *European Business Review*, Vol. 28 No. 2, pp. 208-224.

Monecke, A. and Leisch, F. (2012), "semPLS: structural equation modeling using partial least squares", *Journal of Statistical Software*, Vol. 48 No. 3, pp. 1-32.

Nitzl, C. (2016), "The use of partial least squares structural equation modelling (PLS-SEM) in management accounting research: Directions for future theory development", *Journal of Accounting Literature*, Vol. 37 No. December, pp. 19-35.

Nitzl, C., Roldán, J.L. and Cepeda, C.G. (2016), "Mediation analysis in partial least squares path modeling: Helping researchers discuss more sophisticated models", *Industrial Management and Data Systems*, Vol. 119 No. 9, pp. 1849-1864.

Olsson, U.H., Foss, T., Troye, S.V. and Howell, R.D. (2000), "The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 7 No. 4, pp. 557-595.

Park, S. and Gupta, S. (2012), "Handling endogenous regressors by joint estimation using copulas", *Marketing Science*, Vol. 31 No. 4, pp. 567-586.

Peng, D.X. and Lai, F. (2012), "Using partial least squares in operations management research: a practical guideline and summary of past research", *Journal of Operations Management*, Vol. 30 No. 6, pp. 467-480.

Petter, S. (2018), "Haters gonna hate": PLS and information systems research", *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, Vol. 49 No. 2, pp. 10-13.

Raithel, S., Sarstedt, M., Scharf, S. and Schwaiger, M. (2012), "On the value relevance of customer satisfaction. Multiple drivers and multiple markets", *Journal of the Academy of Marketing Science*, Vol. 40 No. 4, pp. 509-525.

Ramayah, T., Cheah, J.-H., Chuah, F., Ting, H. and Memon, M.A. (2016), *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using SmartPLS 3.0: An Updated and Practical Guide to Statistical Analysis*, Pearson, Singapore.

Ramsey, J.B. (1969), "Tests for specification errors in classical linear least-squares regression analysis", *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 31 No. 2, pp. 350-371.

Rasoolimanesh, S.M. and Ali, F. (2018), "Editorial: partial least squares (PLS) in hospitality and tourism research", *Journal of Hospitality and Tourism Technology*, Vol. 9 No. 3, pp. 238-248.

Reinartz, W.J., Haenlein, M. and Henseler, J. (2009), "An empirical comparison of the efficacy of covariance-based and variance-based SEM", *International Journal of Research in Marketing*, Vol. 26 No. 4, pp. 332-344.

Richter, N.F., Cepeda Carrión, G., Roldán, J.L. and Ringle, C.M. (2016), "European management research using partial least squares structural equation modeling (PLS-SEM): editorial", *European Management Journal*, Vol. 34 No. 6, pp. 589-597.

Richter, N.F., Sinkovics, R.R., Ringle, C.M. and Schlägel, C.M. (2015), "A critical look at the use of SEM in international business research", *International Marketing Review*, Vol. 33 No. 3, pp. 376-404.

Rigdon, E.E. (2012), “Rethinking partial least squares path modeling: in praise of simple methods”, *Long Range Planning*, Vol. 45 Nos 5/6, pp. 341-358.

Rigdon, E.E. (2013), “Partial least squares path modeling”, in Hancock, G.R. and Mueller, R.O. (Eds), *Structural Equation Modeling. A Second Course*, 2 ed. Information Age Publishing, Charlotte NC, pp. 81-116.

Rigdon, E.E. (2014a), “Comment on improper use of endogenous formative variables”, *Journal of Business Research*, Vol. 67 No. 1, pp. 2800-2802.

Rigdon, E.E. (2014b), “Rethinking partial least squares path modeling: breaking chains and forging ahead”, *Long Range Planning*, Vol. 47 No. 3, pp. 161-167.

Rigdon, E.E. (2016), “Choosing PLS path modeling as analytical method in european management research: a realist perspective”, *European Management Journal*, Vol. 34 No. 6, pp. 598-605.

Rigdon, E.E., Sarstedt, M. and Ringle, C.M. (2017), “On comparing results from CB-SEM and PLS-SEM. Five perspectives and five recommendations”, *Marketing Zfp*, Vol. 39 No. 3, pp. 4-16.

Ringle, C.M. and Sarstedt, M. (2016), “Gain more insight from your PLS-SEM results: the Importance-Performance map analysis”, *Industrial Management and Data Systems*, Vol. 116 No. 9, pp. 1865-1886.

Ringle, C.M., Sarstedt, M., Mitchell, R. and Gudergan, S.P. (2019), “Partial least squares structural equation modeling in HRM research”, *The International Journal of Human Resource Management*, Forthcoming.

Ringle, C.M., Sarstedt, M. and Mooi, E.A. (2010), “Response-based segmentation using finite mixture partial least squares: theoretical foundations and an application to american customer satisfaction index data”, *Annals of Information Systems*, Vol. 8, pp. 19-49.

Ringle, C.M., Sarstedt, M. and Straub, D.W. (2012), “A critical look at the use of PLS-SEM in MIS quarterly”, *MIS Quarterly*, Vol. 36 No. 1, pp. iii-xiv.

Ringle, C.M., Wende, S. and Becker, J.-M. (2015), *SmartPLS 3*, SmartPLS, Bönningstedt.

Ringle, C.M., Wende, S. and Will, A. (2005), *SmartPLS 2*, SmartPLS, Hamburg.

Roldán, J.L. and Sánchez-Franco, M.J. (2012), “Variance-based structural equation modeling: guidelines for using partial least squares in information systems research”, in Mora, M., Gelman, O., Steenkamp, AL, *et al.* (Eds), *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems*, IGI Global, Hershey, PA, pp. 193-221.

Sarstedt, M. and Mooi, E.A. (2019), *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics*, Springer, Heidelberg.

Sarstedt, M., Ringle, C.M. and Hair, J.F. (2017a), “Partial least squares structural equation modeling”, in Homburg, C., Klarmann, M. and Vomberg, A. (Eds), *Handbook of Market Research*, Springer, Heidelberg.

Sarstedt, M., Ringle, C.M. and Hair, J.F. (2017b), “Treating unobserved heterogeneity in PLS-SEM: a multi-method approach”, in Noonan, R. and Latan, H. (Eds), *Partial Least Squares Structural Equation Modeling: Basic Concepts, Methodological Issues and Applications*, Springer International Publishing, Cham, pp. 197-217.

Sarstedt, M., Becker, J.-M., Ringle, C.M. and Schwaiger, M. (2011), “Uncovering and treating unobserved heterogeneity with FIMIX-PLS: which model selection criterion provides an appropriate number of segments?”, *Schmalenbach Business Review*, Vol. 63 No. 1, pp. 34-62.

Sarstedt, M., Bengart, P., Shaltoni, A.M. and Lehmann, S. (2018), “The use of sampling methods in advertising research: A gap between theory and practice”, *International Journal of Advertising*, Vol. 37 No. 4, pp. 650-663.

Sarstedt, M., Diamantopoulos, A., Salzberger, T. and Baumgartner, P. (2016a), “Selecting single items to measure doubly-concrete constructs: a cautionary tale”, *Journal of Business Research*, Vol. 69 No. 8, pp. 3159-3167.

Sarstedt, M., Ringle, C.M., Henseler, J. and Hair, J.F. (2014), “On the emancipation of PLS-SEM: a commentary on Rigdon (2012)”, *Long Range Planning*, Vol. 47 No. 3, pp. 154-160.

Sarstedt, M., Hair, J.F., Ringle, C.M., Thiele, K.O. and Gudergan, S.P. (2016b), “Estimation issues with PLS and CBSEM: where the bias lies!”, *Journal of Business Research*, Vol. 69 No. 10, pp. 3998-4010.

Sarstedt, M., Ringle, C.M., Cheah, J.-H., Ting, H., Moisescu, O.I. and Radomir, L. (2019), Structural model robustness checks in PLS-SEM, *Tourism Economics*, Forthcoming.

Schwarz, G. (1978), “Estimating the dimensions of a model”, *The Annals of Statistics*, Vol. 6 No. 2, pp. 461-464.

Sharma, P.N., Sarstedt, M., Shmueli, G., Kim, K.H. and Thiele, K.O. (2019a), “PLS-Based model selection: The role of alternative explanations in information systems research”, *Journal of the Association for Information Systems*, Forthcoming.

Sharma, P.N., Shmueli, G., Sarstedt, M., Danks, S. and Ray, N. (2019b), “Prediction-oriented model selection in partial least squares path modeling”, *Decision Sciences*, Forthcoming.

Shiau, W.-L., Sarstedt, M. and Hair, J.F. (2019), “Editorial: internet research using Partial Least squares Structural equation modeling (PLS-SEM)”, *Internet Research*, Forthcoming.

Shmueli, G. (2010), “To explain or to predict?”, *Statistical Science*, Vol. 25 No. 3, pp. 289-310.

Shmueli, G. and Koppius, O.R. (2011), “Predictive analytics in information systems research”, *MIS Quarterly*, Vol. 35 No. 3, pp. 553-572.

Shmueli, G., Ray, S., Velasquez Estrada, J.M. and Shatla, S.B. (2016), “The elephant in the room: evaluating the predictive performance of PLS models”, *Journal of Business Research*, Vol. 69 No. 10, pp. 4552-4564.

Shmueli, G., Sarstedt, M., Hair, J.F., Cheah, J.-H., Ting, H., Vaithilingam, S. and Ringle, C.M. (2019), “Predictive model assessment in PLS-SEM: guidelines for using PLSpredict”, Working Paper.

Sosik, J.J., Kahai, S.S. and Piovoso, M.J. (2009), “Silver bullet or voodoo statistics? A primer for using the partial least squares data analytic technique in Group and Organization Research”, *Group and Organization Management*, Vol. 34 No. 1, pp. 5-36.

Stone, M. (1974), “Cross-validatory choice and assessment of statistical predictions”, *Journal of the Royal Statistical Society*, Vol. 36 No. 2, pp. 111-147.

Svensson, G., Ferro, C., Høgevold, N., Padin, C., Sosa Varela, J.C. and Sarstedt, M. (2018), “Framing the triple bottom line approach: direct and mediation effects between economic, social and environmental elements”, *Journal of Cleaner Production*, Vol. 197, pp. 972-991.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M. and Lauro, C. (2005), “PLS path modeling”, *Computational Statistics and Data Analysis*, Vol. 48 No. 1, pp. 159-205.

Voorhees, C.M., Brady, M.K., Calantone, R. and Ramirez, E. (2016), “Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies”, *Journal of the Academy of Marketing Science*, Vol. 44 No. 1, pp. 119-134.

Westland, J.C. (2015), “Partial least squares path analysis. Structural equation models: from paths to networks”, Springer International Publishing, Cham, pp. 23-46.

Willaby, H.W., Costa, D.S.J., Burns, B.D., MacCann, C. and Roberts, R.D. (2015), “Testing complex models with small sample sizes: a historical overview and empirical demonstration of what partial least squares (PLS) can offer differential psychology”, *Personality and Individual Differences*, Vol. 84, pp. 73-78.

Wold, H.O.A. (1975), “Path models with latent variables: The NIPALS approach”, in Blalock, H.M., Aganbegian, A., Borodkin, F.M., *et al.* (Eds), *Quantitative Sociology: International*

Perspectives on Mathematical and Statistical Modeling, New York, NY, Academic Press, pp. 307-357.

Wold, H.O.A. (1982), "Soft modeling: the basic design and some extensions", in Jöreskog, K.G. and Wold, H.O.A. (Eds), *Systems under Indirect Observations: Part II*, North-Holland, Amsterdam, pp. 1-54.

Wold, H.O.A. (1985), "Partial least squares", in Kotz, S. and Johnson, N.L. (Eds), *Encyclopedia of Statistical Sciences*, Wiley, New York, NY, pp. 581-591.

Corresponding author
Jeffrey J. Risher can be contacted at: jrisher@uwf.edu