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




## Abstract

**Purpose** – This paper provides a comprehensive, yet concise, overview of the considerations and metrics required for 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 covering established PLS-SEM evaluation criteria, the overview includes new guidelines for applying (1) PLSpredict, a novel approach for assessing a model's out-of-sample prediction, (2) metrics for model comparisons, and (3) 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 results of research, based on the application of PLS-SEM.

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

**Research limitations/implications** – Methodological developments associated with PLS-SEM are rapidly emerging. The metrics reported in this paper are useful for current applications, but scholars need to continuously seek 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 most current and comprehensive summary of the PLS-SEM method and the metrics applied to assess its solutions.

**Keywords** – confirmatory composite analysis, model comparisons, partial least squares, PLSpredict, PLS-SEM, structural equation modeling

**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 sciences journals that used CB-SEM instead of partial least squares SEM (PLS-SEM). By 2015, however, the number of published articles using PLS-SEM increased significantly relative to CB-SEM (Hair et al., 2012b). 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 (Hair et al., 2016a; Ringle et al., 2012), operations management (Peng and Lai, 2012), marketing (Hair et al., 2012b), management accounting (Nitzl, 2016), strategic management (Hair et al., 2012a), hospitality (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) also illustrate PLS-SEM or propose methodological extensions.

The primary appeal of PLS-SEM is that the method enables researchers 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 structure is 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). Finally, 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 in assessing 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). Researchers should routinely apply PLSpredict in their PLS-SEM studies to assess the predictive power (Shmueli et al., 2019), 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). Finally, we elaborate on using PLS-SEM for a confirmatory composite analysis (CCA), which allows developing or revising both reflectively and formatively measured construct measures within a nomological net of constructs. Figure 1 illustrates the various aspects that we discuss in the following sections.

INSERT FIGURE 1 HERE

## PRELIMINARY CONSIDERATIONS

The Swedish econometrician Herman O. A. Wold (1975; 1982; 1985) developed the statistical underpinnings of the PLS-SEM method, which was initially known and is sometimes still referred to as PLS path modeling (Hair et al., 2011). PLS-SEM estimates partial model structures, which are defined in a path model, by combining principal components analysis with ordinary least squares regressions (Mateos-Aparicio, 2011). The 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, since it accounts for the total variance and uses the total variance to estimate parameters (Rigdon et al., 2017a; Hair et al., 2017c).

In the past decade, there has been considerable debate about which situations are more or less appropriate for using PLS-SEM (e.g., Goodhue et al., 2012; Marcoulides et al., 2012; Marcoulides and Saunders, 2006; Rigdon, 2014a; Henseler et al., 2014; Khan et al., 2018). In the following sections, we summarize several initial considerations when choosing to use PLS-SEM (e.g., Hair et al., 2013). When helpful, we compare the differences between CB-SEM and PLS-SEM (e.g., Marcoulides and Chin, 2013; Rigdon, 2016). In doing so, we note that recent research has moved beyond the CB-SEM vs. PLS-SEM debate (e.g., Rigdon et al., 2017b; Rigdon, 2012), thereby establishing PLS-SEM as a distinct method for analyzing composite-based path models.

Nevertheless, applied research is still frequently confronted with the choice between the two SEM methods. Against this background, Table 1 provides an overview of points to consider when deciding whether PLS is an appropriate SEM method for a study.

INSERT TABLE 1 HERE



**Sample Size**

PLS-SEM obtains 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., 2017c). 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 obtains solutions when methods such as CB-SEM develop inadmissible solutions or do not converge at all with complex models and small sample sizes, regardless of whether the data originate from a common or composite model population. Hair et al. (2013) indicate, however, that certain scholars have falsely and misleadingly taken advantage of these characteristics to obtain solutions using extremely small sample sizes, even when the population is large and accessible without much effort. This practice has unfortunately tarnished the reputation of PLS-SEM to some extent (also see Marcoulides et al., 2009). Like other multivariate methods, PLS-SEM is not capable of turning a bad (e.g., non-representative) sample into a proper one to obtain valid model estimations.

PLS-SEM can, however, 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. In order 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 (e.g. 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 datasets, 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 (e.g., Hair et al., 2012b; Nitzl, 2016; do Valle and Assaker, 2015). While this is clearly an advantage of using PLS-SEM in social sciences studies, which almost always rely on nonnormal data, on its own, it is not sufficient justification.

Scholars have noted that maximum likelihood estimation with CB-SEM is robust against violations of normality (e.g., 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 dataset is limited, CB-SEM can produce abnormal results when data are nonormal (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 bootstrap distributions. The use of the bias-corrected and accelerated (BCa) bootstrapping routine minimizes this issue to some extent, since it adjusts the confidence intervals for skewness (Efron, 1987). Choosing to only use PLS-SEM on data distribution 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 CB-SEM analyses that are strictly confirmatory in nature. More precisely, secondary data are primarily used in exploratory research to propose causal relationships in situations which have little clearly defined theory (Hair, Hollingsworth, et al., 2017). 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 parameter requirements of CB-SEM (Hair et al., 2014). Thus, PLS-SEM is preferred 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: 29) notes, "soft modeling is primarily designed for research contexts that are simultaneously data-rich and theory-skeletal." Furthermore, the increasing popularity of a 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 fit the prediction-oriented PLS-SEM approach like hand in glove.

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 In addition, when using secondary data researchers do not have an 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 use of formative measures (Hair et al., 2017d). Since the PLS-SEM algorithm is based on ordinary least squares regression, the method allows unrestricted use of single-item, reflective, or 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., 2017). For instance, Ittner et al. (1997) operationalized strategy with four indicators: (1) the ratio of research and development to sales, (2) the market-to-book ratio, (3) the ratio of employees to sales, and (4) 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 model when using formative measures, since using 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., 2017c). 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 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 very appropriate for exploratory research, but the method is also useful for confirmatory research (Hair et al., 2017a).

## Goodness-of-fit

Unlike CB-SEM, which strongly relies on the concept of model fit, this is much less the case with PLS-SEM (Hair et al., 2019). This has led some researchers to incorrectly conclude that PLS-SEM is not useful for theory testing and confirmation (e.g., Westland, 2015). A couple of methodologists have attempted to develop 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 thus far. Therefore, any thresholds (guidelines) advocated in the literature should be considered as very tentative. Second, since 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—is 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 (e.g., 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 PLS-SEM RESULTS

The first step in evaluating PLS-SEM results involves examining the measurement models, which differs for reflective and formative constructs. If the measurement models meet 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. But 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, since they indicate that the construct explains more than 50 percent 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 0.70 and 0.90 range from "satisfactory to good." But values of 0.95 and higher are problematic, since 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 since the items are unweighted. In contrast, with composite reliability, the items are weighted based on the construct indicators' individual loadings and 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 $\rho_A$ as an approximately exact measure of construct reliability, which usually lies between Cronbach's alpha and the composite reliability. Hence, $\rho_A$ 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% 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% confidence interval of the construct reliability is lower than 0.95). In order 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 is to assess the convergent validity of each construct measure. Convergent validity is the extent to which the construct converges in order 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. In order to calculate the AVE, one has to square the loading of each indicator on a construct and compute the mean value. The minimum acceptable AVE is 0.50 or higher—an AVE of 0.50 or higher indicates that the construct explains 50 percent or more of the variance of the items that make up the construct.

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 in 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 ratio (HTMT) of the correlations (also see Voorhees et al., 2016). The HTMT is defined as the mean value of the item correlations across constructs (i.e., the heterotrait-heteromethod correlations) relative to the (geometric) mean of the average correlations for the items measuring the same construct (i.e., the monotrait-heteromethod correlations). 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, such as 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% confidence interval of HTMT is lower than 0.9 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 (1) convergent validity, (2) indicator collinearity, and (3) 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. In order to execute this procedure for determining convergent validity, researchers must plan ahead in the research design stage by including an alternative reflectively measured indicator of the formatively measured construct in their questionnaire. Cheah et al. (2019) 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). But a multi-item measure is acceptable and perhaps better. When the model is based on secondary data, a variable measuring a similar concept would be used (Houston, 2004), but it would seldom be measured reflectively. 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. When VIF values are higher, the level of collinearity is greater. VIF values of 5 or above indicate collinearity issues among the predictor 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 ot construct bootstrap-based confidence intervals (also see 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. But 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 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 (e.g., Cenfetelli and Bassellier, 2009). Second, indicators should seldom be removed from formative measurement models, since 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 (e.g., 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 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 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 $Q^2$, as well as 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), assuming the sample size is large enough.

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 exogenous constructs 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 to 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., 2017b).

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, the $R^2$ values of 0.75, 0.50, and 0.25 can be considered substantial, moderate, and weak (Henseler et al., 2009; Hair et al., 2011). However, 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, as, for example, in predicting stock returns (e.g., Raithel et al., 2012). In addition, 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, $R^2$ 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, thereby indicating that the model overfits the data. That is, the model is too complex, thereby fitting the random noise inherent in the sample rather than reflecting the overall population. The same model would likely not fit if used on another sample drawn from the population (Sharma et al. 2019a). When measuring a concept that is inherently predictable, such as physical processes, $R^2$ values of 0.9 might not be surprising. But the same $R^2$ values in a model that predicts human attitudes, perceptions, and intentions could likely indicate 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. Otherwise (i.e., 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, for example, of 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 therefore 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 0.5 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, since 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. The PLSpredict procedure generates the holdout sample-based predictions in PLS-SEM and is an option in standard 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 dataset 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. When the sample is too small to use k=10, a smaller k value can be used. Also, two other criteria to assess out-of-sample prediction without using a holdout sample are available – the BIC and GM (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 generally running 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 dataset), PLSpredict should be run only once (i.e., without repetitions).

For the assessment of a model's predictive power when using PLSpredict, 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. Since 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 $Q^2_{\text{predict}}$ statistic should be evaluated first to verify that 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. But 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 therefore 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). In 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 produces 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 greater 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. Researchers can also 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 to 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) Bayesian information criterion (BIC) and Geweke and Meese's (1981) criterion (GM) achieve a sound tradeoff 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 samples that are too small to permit dividing into analysis and hold out 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.

Table 2 summarizes the metrics that need to be applied when interpreting and reporting PLS-SEM results.

INSERT TABLE 2 HERE

**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.

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 vs. 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 are 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 zero, 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 potential (1) nonlinear effects, (2) endogeneity, and (3) unobserved heterogeneity. First, to test whether or not relationships are nonlinear, researchers can run Ramsey's (1969) RESET 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 only, rather than when the focus is on PLS-SEM's causal-predictive character.

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 dataset is very likely to produce misleading results (Becker et al., 2013). Hence, any PLS-SEM analysis should include a routine check for unobserved heterogeneity to ascertain whether or not the analysis of the entire dataset 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., 2016b; Matthews et al., 2016). If heterogeneity is present at a critical level, the next step involves running the PLS prediction-oriented segmentation (Becker et al., 2013) procedure on the data to disclose the segment structure. Finally, researchers 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 PLS-SEM 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) when appropriate, can identify the analysis with further particularized findings, conclusions, and implications.

**PLS-SEM and confirmation of measurement scales**

Confirmatory factor analysis (CFA) has historically been used to develop and improve reflectively measured constructs based on the domain sampling model (Hair et al., 2019). Compared to CFA, CCA is a recently proposed alternative approach that offers several advantages. CCA is a series of steps executed with PLS-SEM to confirm both reflective and formative measurement models of established measures that are being updated or adapted to a different context. Note that CCA is also useful for developing new measures.

Before executing a CCA, operational definitions of the multi-item construct must be confirmed, including whether the appropriate measurement model is reflective or formative, since the process for the two types of measurement differs considerably. These initial steps are followed by literature reviews and qualitative research with expert panels to assess face validity and reduce the initial list of items. Pilot testing for refinement and purification of the items prepares the researcher for executing a CCA.

CCA differs from CFA in that the statistical objective is to maximize variance extracted from the exogenous variables, but in doing so to facilitate prediction of the endogenous constructs, and confirmation of the measurement models. That is, CCA enables researchers to develop and validate measures within a nomological network. The method is an extension of principal components analysis since it is composite-based and, therefore, produces composite scores that are weighted sums of indicators and can be used in follow-up analyses. The resulting composites are correlated, however, as they would be in an oblique rotation with an exploratory factor analysis, and include variance that maximizes prediction of the endogenous constructs. Note that the composite correlations from the oblique rotation seldom result in problems with multicollinearity, but this issue should always be examined (Hair et al., 2017a).

To summarize the CCA execution process, first assess the standard PLS measurement model criteria for item reliability, item loadings, internal consistency reliability, convergent validity, and discriminant validity. If these metrics meet the recommended guidelines, the next step is to assess nomological validity by comparing the constructs included in the CCA to another construct within the nomological net. The third and final CCA step is to assess the criterion or preferably the predictive validity of the CCA constructs based on the structural model results.

To achieve measurement confirmation objectives in developing or adapting multi-item measures, researchers could use either CFA or CCA. The results are different, however, and researchers need to understand the implications of the distinct outcomes to make informed decisions. CCA and CFA can both be used to improve item and scale reliability, identify and provide an indication of items that need to be revised or in some instances eliminated for content validity, facilitate achieving convergent validity and discriminant validity, and to remove error variance. Compared to CFA, CCA has several benefits, which are as follows: (1) the number of items retained to measure constructs is higher with CCA, thereby improving construct validity, (2) determinant construct scores are available (Rigdon et al., 2019), and (3) CCA can be applied to formative measurement models.

## 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 (e.g., 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 source to rely 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. We also summarize an emerging role for confirmatory composite analysis (CCA), which is a alternative to the use of confirmatory factor analysis (CFA) in the development, adaptation and confirmation of measurement scales. 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 MI and Rönkkö M. (2018) Statistical Inference with PLSc Using Bootstrap Confidence Intervals. *MIS Quarterly* 42(3): 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 & Control* 28(11-12): 1011-1021.

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

Ali F, Rasoolimanesh SM, Sarstedt M, et al. (2018b) An Assessment of the Use of Partial Least Squares Structural Equation Modeling (PLS-SEM) in Hospitality Research. *International Journal of Contemporary Hospitality Management* 30(1): 514-538.

Avkiran NK and Ringle CM. (2018) Partial Least Squares Structural Equation Modeling: Recent Advances in Banking and Finance. Cham: Springer International Publishing.

Bascle G. (2008) Controlling for Endogeneity with Instrumental Variables in Strategic Management Research. *Strategic Organization* 6(3): 285-327.

Becker J-M, Rai A, Ringle CM, et al. (2013) Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats. *MIS Quarterly* 37(3): 665-694.

Becker J-M, Ringle CM and Sarstedt M. (2018) Estimating Moderating Effects in PLS-SEM and PLSc-SEM: Interaction Term Generation*Data Treatment. *Journal of Applied Structural Equation Modeling* 2(2): 1-21.

Bollen KA and Ting K-f. (2000) A Tetrad Test for Causal Indicators. *Psychological Methods* 5(1): 3-22.

Boomsma A and Hoogland JJ. (2001) The Robustness of LISREL Modeling Revisited. In: Cudeck R, du Toit S and Sörbom D (eds) *Structural Equation Modeling: Present and Future.* Chicago: Scientific Software International, 139-168.

Cenfetelli RT and Bassellier G. (2009) Interpretation of Formative Measurement in Information Systems Research. *MIS Quarterly* 33(4): 689-708.
Cheah J-H, Sarstedt M, Ringle CM, et al. (2019) 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* forthcoming.
Chin WW. (1998) The Partial Least Squares Approach to Structural Equation Modeling. In: Marcoulides GA (ed) *Modern Methods for Business Research.* Mahwah: Erlbaum, 295-358.
Chin WW. (2003) PLS-Graph 3.0. Houston: Soft Modeling Inc.
Chin WW. (2010) How to Write Up and Report PLS Analyses. 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, vol. II).* Heidelberg, Dordrecht, London, New York: Springer, 655-690.
Chin WW 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 WW, Henseler J, et al. (eds) *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series, vol. II).* Heidelberg, Dordrecht, London, New York: Springer, 171-193.
Chou C-P, Bentler PM 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* 44(2): 347-357.
Cochran WG. (1977) *Sampling Techniques,* New York, NY: Wiley.
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 SM and Cobanoglu C (eds) *Applying Partial Least Squares in Tourism and Hospitality Research.* Bingley: Emerald, 35-52.
Diamantopoulos A, Sarstedt M, Fuchs C, et al. (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* 40(3): 434-449.
Diamantopoulos A and Winklhofer HM. (2001) Index Construction with Formative Indicators: An Alternative to Scale Development *Journal of Marketing Research* 38(2): 269-277.
Dijkstra TK and Henseler J. (2015) Consistent Partial Least Squares Path Modeling. *MIS Quarterly* 39(2): 297-316.
do Valle PO and Assaker G. (2015) Using Partial Least Squares Structural Equation Modeling in Tourism Research: A Review of Past Research and Recommendations for Future Applications. *Journal of Travel Research* in press.
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.* Cham: Springer International Publishing, 153-172.
Drolet AL and Morrison DG. (2001) Do We Really Need Multiple-Item Measures in Service Research? *Journal of Service Research* 3(3): 196-204.
Efron B. (1987) Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association* 82(397): 171-185.

Fornell CG and Bookstein FL. (1982) Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. *Journal of Marketing Research* 19(4): 440-452.
Fornell CG and Larcker DF. (1981) Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research* 18(1): 39-50.
Franke GR and Sarstedt M. (2019) Heuristics Versus Statistics in Discriminant Validity Testing: A Comparison of Four Procedures. *Internet Research* forthcoming.
Garson GD. (2016) *Partial Least Squares Regression and Structural Equation Models,* Asheboro: Statistical Associates.
Geisser S. (1974) A Predictive Approach to the Random Effects Model. *Biometrika* 61(1): 101-107.
Geweke J and Meese R. (1981) Estimating Regression Models of Finite but Unknown Order. *Journal of Econometrics* 16(1): 162.
Goodhue DL, Lewis W and Thompson R. (2012) Does PLS Have Advantages for Small Sample Size or Non-Normal Data? *MIS Quarterly* 36 (3): 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 WW, Henseler J, et al. (eds) *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series, vol. II).* Heidelberg, Dordrecht, London, New York: Springer, 691-711.
Gudergan SP, Ringle CM, Wende S, et al. (2008) Confirmatory Tetrad Analysis in PLS Path Modeling. *Journal of Business Research* 61(12): 1238-1249.
Hahn C, Johnson MD, Herrmann A, et al. (2002) Capturing Customer Heterogeneity Using a Finite Mixture PLS Approach. *Schmalenbach Business Review* 54(3): 243-269.
Hair JF, Hollingsworth CL, Randolph AB, et al. (2016a) An Updated and Expanded Assessment of PLS-SEM in Information Systems Research. *Industrial Management & Data Systems* in press.
Hair JF, Hult GTM, Ringle CM, et al. (2017a) *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM),* Thousand Oaks, CA: Sage.
Hair JF, Hult GTM, Ringle CM, et al. (2017b) *A Primer on Partial Least Squares Structural Equations Modeling (PLS-SEM),* Thousand Oaks: SAGE.
Hair JF, Hult GTM, Ringle CM, et al. (2017c) Mirror, Mirror on the Wall: A Comparative Evaluation of Composite-based Structural Equation Modeling Methods *Journal of the Academy of Marketing Science* 45(5): 616-632.
Hair JF, Ringle CM and Sarstedt M. (2011) PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice* 19(2): 139-151.
Hair JF, Ringle CM and Sarstedt M. (2013) Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. *Long Range Planning* 46(1-2): 1-12.
Hair JF, Sarstedt M, Hopkins L, et al. (2014) Partial Least Squares Structural Equation Modeling (PLS-SEM): An Emerging Tool in Business Research. *European Business Review* 26(2): 106-121.
Hair JF, Sarstedt M, Matthews L, et al. (2016b) Identifying and Treating Unobserved Heterogeneity with FIMIX-PLS: Part I - Method. *European Business Review* 28(1): 63-76.
Hair JF, Sarstedt M, Pieper TM, et al. (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* 45(5-6): 320-340.
Hair JF, Sarstedt M and Ringle CM. (2019) Rethinking Some of the Rethinking of Partial Least

Squares. *European Journal of Marketing* forthcoming.
Hair JF, Sarstedt M, Ringle CM, et al. (2018) *Advanced Issues in Partial Least Squares Structural Equation Modeling (PLS-SEM),* Thousand Oaks, CA: Sage.
Hair JF, Sarstedt M, Ringle CM, 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* 40(3): 414-433.
Henseler J, Dijkstra TK, Sarstedt M, et al. (2014) Common Beliefs and Reality about Partial Least Squares: Comments on Rönkkö & Evermann (2013). *Organizational Research Methods* 17(2): 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, vol. II).* Heidelberg, Dordrecht, London, New York: Springer, 713-735.
Henseler J, Hubona GS and Ray PA. (2016a) Using PLS Path Modeling in New Technology Research: Updated Guidelines. *Industrial Management & Data Systems* 116(1): 1-19.
Henseler J, Hubona GS and Ray PA. (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.* Heidelberg: Springer, 19-39.
Henseler J, Ringle CM and Sarstedt M. (2015) A New Criterion for Assessing Discriminant Validity in Variance-based Structural Equation Modeling. *Journal of the Academy of Marketing Science* 43(1): 115-135.
Henseler J, Ringle CM and Sarstedt M. (2016b) Testing Measurement Invariance of Composites Using Partial Least Squares. *International Marketing Review* 33(3): 405-431.
Henseler J, Ringle CM and Sinkovics RR. (2009) The Use of Partial Least Squares Path Modeling in International Marketing. In: Sinkovics RR and Ghauri PN (eds) *Advances in International Marketing.* Bingley: Emerald 277-320.
Henseler J and Sarstedt M. (2013) Goodness-of-Fit Indices for Partial Least Squares Path Modeling. *Computational Statistics* 28(2): 565-580.
Houston MB. (2004) Assessing the Validity of Secondary Data Proxies for Marketing Constructs. *Journal of Business Research* 57(2): 154-161.
Hult GTM, Hair JF, Proksch D, et al. (2018) Addressing Endogeneity in International Marketing Applications of Partial Least Squares Structural Equation Modeling. *Journal of International Marketing* 26(3): 1-21.
Jöreskog KG. (1971) Simultaneous factor analysis in several populations. *Psychometrika* 36(4): 409-426.
Jöreskog KG. (1973) A General Method for Estimating a Linear Structural Equation System. In: Goldberger AS and Duncan OD (eds) *Structural Equation Models in the Social Sciences.* New York: Seminar Press, 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* 21(4): 259–272.
Khan GF, Sarstedt M, Shiau W-L, et al. (2018) 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* 28(1): 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 SM and Cobanoglu C (eds) *Applying Partial Least Squares in Tourism and Hospitality Research.* Bingley: Emerald, 53-84.
Lohmöller J-B. (1989) *Latent Variable Path Modeling with Partial Least Squares,* Heidelberg: Physica.
Marcoulides GA and Chin WW. (2013) You Write, but Others Read: Common Methodological Misunderstandings in PLS and Related Methods. In: Abdi H, Chin WW, Esposito Vinzi V, et al. (eds) *New Perspectives in Partial Least Squares and Related Methods.* Springer New York, 31-64.
Marcoulides GA, Chin WW and Saunders C. (2009) Foreword: A Critical Look at Partial Least Squares Modeling. *MIS Quarterly* 33(1): 171-175.
Marcoulides GA, Chin WW and Saunders C. (2012) When Imprecise Statistical Statements Become Problematic: A Response to Goodhue, Lewis, and Thompson. *MIS Quarterly* 36(3): 717-728.
Marcoulides GA and Saunders C. (2006) PLS: A Silver Bullet? *MIS Quarterly* 30(2): III-IX.
Mateos-Aparicio G. (2011) Partial Least Squares (PLS) Methods: Origins, Evolution, and Application to Social Sciences. *Communications in Statistics - Theory and Methods* 40(13): 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.* Heidelberg: Springer, 219-243.
Matthews L, Sarstedt M, Hair JF, et al. (2016) Identifying and Treating Unobserved Heterogeneity with FIMIX-PLS: Part II – A Case Study. *European Business Review* 28(2): 208-224.
Monecke A and Leisch F. (2012) semPLS: Structural Equation Modeling Using Partial Least Squares. *Journal of Statistical Software* 48(3): 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* 37(December): 19-35.
Nitzl C, Roldán JL and Cepeda Carrión G. (2016) Mediation Analysis in Partial Least Squares Path Modeling: Helping Researchers Discuss More Sophisticated Models. *Industrial Management & Data Systems* 119(9): 1849-1864.
Olsson UH, Foss T, Troye SV, et al. (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* 7(4): 557-595.
Park S and Gupta S. (2012) Handling Endogenous Regressors by Joint Estimation Using Copulas. *Marketing Science* 31(4): 567-586.
Peng DX and Lai F. (2012) Using Partial Least Squares in Operations Management Research: A Practical Guideline and Summary of Past Research. *Journal of Operations Management* 30(6): 467–480.
Petter S. (2018) "Haters Gonna Hate": PLS and Information Systems Research. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 49(2): 10-13
Raithel S, Sarstedt M, Scharf S, et al. (2012) On the value relevance of customer satisfaction. Multiple drivers and multiple markets. *Journal of the Academy of Marketing Science* 40(4): 509-525.
Ramayah T, Cheah J-H, Chuah F, et al. (2016) *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using SmartPLS 3.0: An Updated and Practical Guide to*

*Statistical Analysis,* Singapore et al. : Pearson.
Ramsey JB. (1969) Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* 31(2): 350-371.
Rasoolimanesh SM and Ali F. (2018) Editorial: Partial Least Squares (PLS) in Hospitality and Tourism Research. *Journal of Hospitality and Tourism Technology* forthcoming.
Reinartz WJ, 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* 26(4): 332-344.
Richter NF, Cepeda Carrión G, Roldán JL, et al. (2016) European Management Research Using Partial Least Squares Structural Equation Modeling (PLS-SEM): Editorial. *European Management Journal* 34(6): 589-597.
Richter NF, Sinkovics RR, Ringle CM, et al. (2015) A Critical Look at the Use of SEM in International Business Research *International Marketing Review* 33(3): 376-404.
Rigdon EE. (2012) Rethinking Partial Least Squares Path Modeling: In Praise of Simple Methods. *Long Range Planning* 45(5-6): 341-358.
Rigdon EE. (2013) Partial Least Squares Path Modeling. In: Hancock GR and Mueller RO (eds) *Structural Equation Modeling. A Second Course.* 2 ed. Charlotte NC: Information Age Publishing, 81-116.
Rigdon EE. (2014a) Comment on "Improper Use of Endogenous Formative Variables". *Journal of Business Research* 67(1): 2800-2802.
Rigdon EE. (2014b) Rethinking Partial Least Squares Path Modeling: Breaking Chains and Forging Ahead. *Long Range Planning* 47(3): 161-167.
Rigdon EE. (2016) Choosing PLS Path Modeling as Analytical Method in European Management Research: A Realist Perspective. *European Management Journal* 34(6): 598-605.
Rigdon EE, Becker J-M and Sarstedt M. (2017a) Equating Unobserved Conceptual Variables and Common Factors in Structural Equation Models. *Working Paper.*
Rigdon EE, Becker J-M and Sarstedt M. (2019) Factor Indeterminacy as Metrological Uncertainty: Implications for Advancing Psychological Measurement,. *Multivariate Behavioral Research* forthcoming.
Rigdon EE, Sarstedt M and Ringle CM. (2017b) On Comparing Results from CB-SEM and PLS-SEM. Five Perspectives and Five Recommendations. *Marketing ZFP* 39(3): 4-16.
Ringle CM and Sarstedt M. (2016) Gain More Insight from Your PLS-SEM Results: The Importance-Performance Map Analysis. *Industrial Management & Data Systems* 116(9): 1865-1886.
Ringle CM, Sarstedt M, Mitchell R, et al. (2019) Partial Least Squares Structural Equation Modeling in HRM Research. *The International Journal of Human Resource Management* forthcoming.
Ringle CM, Sarstedt M and Mooi EA. (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* 8: 19-49.
Ringle CM, Sarstedt M and Straub DW. (2012) A Critical Look at the Use of PLS-SEM in MIS Quarterly. *MIS Quarterly* 36(1): iii-xiv.
Ringle CM, Wende S and Becker J-M. (2015) SmartPLS 3. Bönningstedt: SmartPLS.
Ringle CM, Wende S and Will A. (2005) SmartPLS 2. Hamburg: SmartPLS.
Roldán JL and Sánchez-Franco MJ. (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.* Hershey, PA: IGI Global, 193-221.

Sarstedt M, Becker J-M, Ringle CM, et al. (2011) Uncovering and Treating Unobserved Heterogeneity with FIMIX-PLS: Which Model Selection Criterion Provides an Appropriate Number of Segments? *Schmalenbach Business Review* 63(1): 34-62.

Sarstedt M, Bengart P, Shaltoni AM, et al. (2018) The Use of Sampling Methods in Advertising Research: A Gap Between Theory and Practice. *International Journal of Advertising* 37(4): 650-663.

Sarstedt M, Diamantopoulos A, Salzberger T, et al. (2016a) Selecting Single Items to Measure Doubly-Concrete Constructs: A Cautionary Tale. *Journal of Business Research* 69(8): 3159-3167.

Sarstedt M, Hair JF, Ringle CM, et al. (2016b) Estimation Issues with PLS and CBSEM: Where the Bias Lies! *Journal of Business Research* 69(10): 3998-4010.

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

Sarstedt M, Ringle CM, Cheah J-H, et al. (2019) Structural Model Robustness Checks in PLS-SEM. *Tourism Economics* forthcoming.

Sarstedt M, Ringle CM and Hair JF. (2017a) Partial Least Squares Structural Equation Modeling. In: Homburg C, Klarmann M and Vomberg A (eds) *Handbook of Market Research.* Heidelberg: Springer.

Sarstedt M, Ringle CM and Hair JF. (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.* Heidelberg: Springer, 197-217.

Sarstedt M, Ringle CM, Henseler J, et al. (2014) On the Emancipation of PLS-SEM: A Commentary on Rigdon (2012). *Long Range Planning* 47(3): 154-160.

Schwarz G. (1978) Estimating the Dimensions of a Model. *Annals of Statistics* 6(2): 461-464.

Sharma PN, Sarstedt M, Shmueli G, et al. (2019a) PLS-Based Model Selection: The Role of Alternative Explanations in Information Systems Research. *Journal of the Association for Information Systems* in press.

Sharma PN, Shmueli G, Sarstedt M, et al. (2019b) Prediction-oriented Model Selection in Partial Least Squares Path Modeling. *Decision Sciences* in press.

Shiau W-L, Sarstedt M and Hair JF. (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* 25(3): 289-310.

Shmueli G and Koppius OR. (2011) Predictive Analytics in Information Systems Research. *MIS Quarterly* 35(3): 553-572.

Shmueli G, Ray S, Velasquez Estrada JM, et al. (2016) The Elephant in the Room: Evaluating the Predictive Performance of PLS Models. *Journal of Business Research* 69(10): 4552-4564.

Shmueli G, Sarstedt M, Cheah J, et al. (2019) Predictive Model Assessment in PLS-SEM: Guidelines for Using PLSpredict. *European Journal of Marketing* forthcoming.

Sosik JJ, Kahai SS and Piovoso MJ. (2009) Silver bullet or voodoo statistics? A primer for using the partial least squares data analytic technique in group and organization research. *Group & Organization Management* 34(1): 5-36.

Stone M. (1974) Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society* 36(2): 111-147.

Svensson G, Ferro C, Høgevold N, et al. (2018) Framing the Triple Bottom Line Approach:

Direct and Mediation Effects Between Economic, Social and Environmental Elements. *Journal of Cleaner Production* 197(Part 1): 972-991.

Tenenhaus M, Esposito Vinzi V, Chatelin Y-M, et al. (2005) PLS Path Modeling. *Computational Statistics & Data Analysis* 48(1): 159-205.

Voorhees CM, Brady MK, Calantone R, et al. (2016) Discriminant Validity Testing in Marketing: An Analysis, Causes for Concern, and Proposed Remedies. *Journal of the Academy of Marketing Science* 44(1): 119-134.

Westland JC. (2015) Partial Least Squares Path Analysis. *Structural Equation Models: From Paths to Networks.* Cham: Springer International Publishing, 23-46.

Willaby HW, Costa DSJ, Burns BD, et al. (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* 84: 73-78.

Wold HOA. (1975) Path Models with Latent Variables: The NIPALS Approach. In: Blalock HM, Aganbegian A, Borodkin FM, et al. (eds) *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling.* New York et al.: Academic Press, 307-357.

Wold HOA. (1982) Soft Modeling: The Basic Design and Some Extensions. In: Jöreskog KG and Wold HOA (eds) *Systems Under Indirect Observations: Part II.* Amsterdam: North-Holland, 1-54.

Wold HOA. (1985) Partial Least Squares. In: Kotz S and Johnson NL (eds) *Encyclopedia of Statistical Sciences.* New York: Wiley, 581-591.

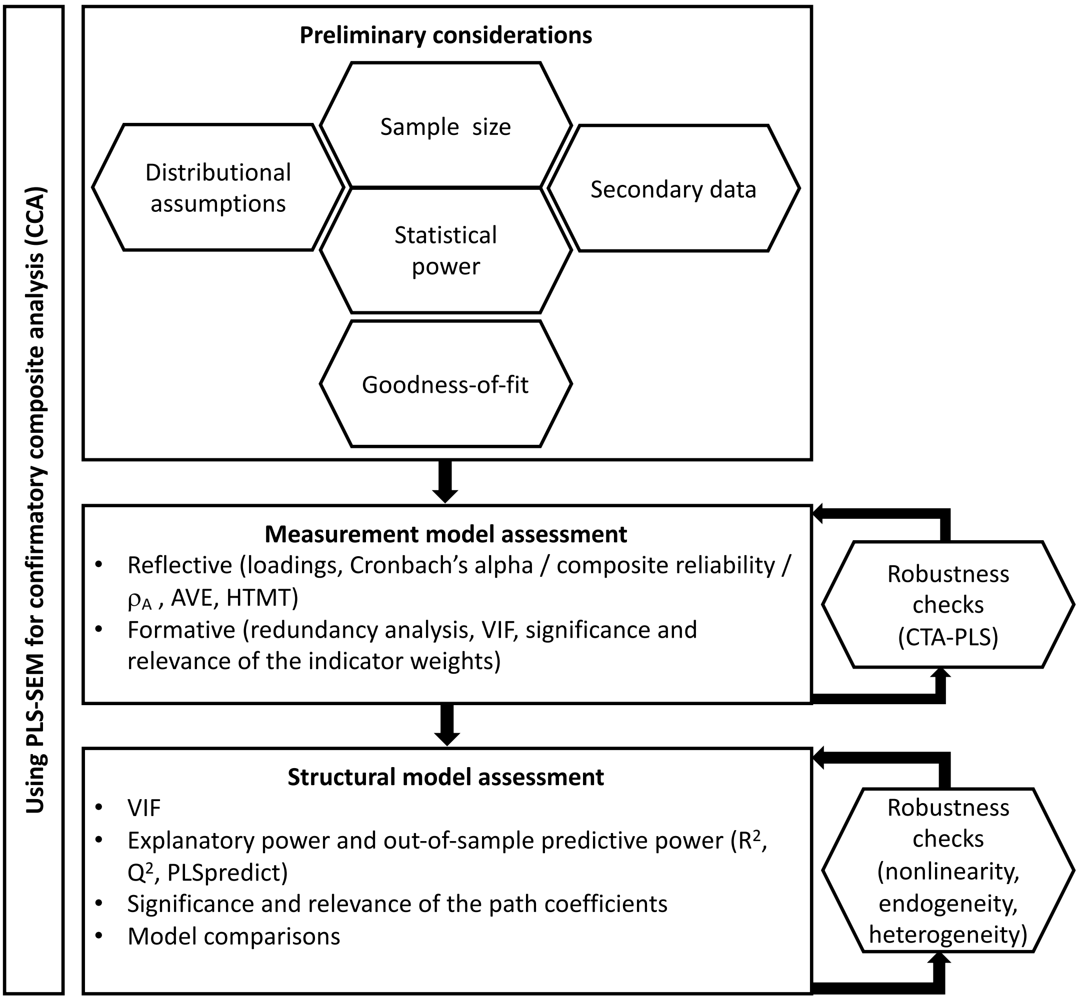


**Figure 1.** Aspects and statistics to consider in a PLS-SEM analysis

| |
|---|
| **Selection of PLS-SEM** |
| • When the analysis in concerned with **testing a theoretical framework** from a **prediction perspective**. |
| • The **structural model is complex** and includes many constructs, indicators, and/or model relationships. |
| • The research objective is better **understanding 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**. |
| • The research consists of financial ratios or similar types of **data artifacts**. |
| • 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. |
| • When research requires **latent variable scores** for follow-up analyses. |

**Table 1.** Conditions that favor the use of PLS-SEM

| | |
|---|---|
| **Reflective measurement models** | |
| Reflective indicator loadings | • ≥ 0.708 |
| Internal consistency reliability | • Cronbach's alpha is the lower bound, the composite reliability is the upper bound for internal consistency reliability. $\rho_A$ 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.<br>• Minimum 0.70 (or 0.60 in exploratory research)<br>• Maximum of 0.95 to avoid indicator redundancy, which would compromise content validity<br>• Recommended 0.80 to 0.90<br>• 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 ≥ 0.50 |
| Discriminant validity | • For conceptually similar constructs: HTMT < 0.90<br>• For conceptually different constructs: HTMT < 0.85<br>• Test if the HTMT is significantly lower than the threshold value |
| **Formative measurement models** | |
| Convergent validity (redundancy analysis) | • ≥ 0.70 correlation |
| Collinearity (VIF) | • Probable (i.e., critical) collinearity issues when VIF ≥ 5<br>• Possible collinearity issues when VIF ≥ 3 – 5<br>• 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 ≥ 0.50 that are statistically significant are considered relevant. |
| **Structural model** | |
| Collinearity (VIF) | • Probable (i.e., critical) collinearity issues when VIF ≥ 5<br>• Possible collinearity issues when VIF ≥ 3 – 5<br>• 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.<br>• Values higher than 0, 0.25, and 0.50 depict small, medium, and large predictive relevance of the PLS-path model |
| PLSpredict | • Set *k*=10, assuming each subgroup meets the minimum required sample size.<br>• Use ten repetitions, assuming the sample size is large enough.<br>• $Q^2_{predict}$ values ≤ 0 indicate that the model does not outperform the most naïve benchmark (i.e., the indicator means from the analysis sample).<br>• Compare the MAE (or the RMSE) value with the LM value of each 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<br>• Endogeneity<br>• Unobserved heterogeneity |

**Table 2.** Guidelines when using PLS-SEM