International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

# PARTIAL LEAST SQUARES (PLS-SEM): A NOTE FOR BEGINNERS

**[1]Mohammed Lawal Yahaya; Zainab Ango Murtala[2] & Henry. N. Onukwube[3]**
[1]Department of Physical Planning and Development, Usmanu Danfodiyo University, Sokoto
[2]Department of Works and Services, Federal Neuro-Psychiatric Hospital Kware, Sokoto
[3]Department of Quantity Surveying, Faculty of Environmental Sciences, University of Lagos, Lagos
**Email:** mohammedlawalyahaya@gmail.com

## ABSTRACT

Smart PLS is among the leading software applications for Partial Least Squares Structural Equation Modeling (PLS-SEM). It has been developed by Ringle, Wende & Will (2005). Since its launch in 2005, the software has gained popularity not only because it is freely available to academics and researchers, but also because it has a user-friendly interface, advanced reporting features and choose formative or reflective models. Although a large number of journal articles on the subject of PLS modeling have been published, the number of instruction materials available for this software is limited especially within the context of environmental studies. This paper is written to resolve this knowledge gap and to help beginners in various field of study to understand how PLS-SEM can be used in construction economics and management research. A step-by- step method was demonstrated for the measurement and structural model evaluation stating the main points in reporting each model. We believe that construction industry scholars would recognize and embraced the use of the statistical methods at their disposal to explore and better understand the phenomena they are researching compared to the first generation tool of analysis employed earlier.
**Keywords:** Bootstrapping; Formative; Measurement Model; Path coefficient; Reflective; Smart PLS; Structural Model.

## INTRODUCTION

First-generation methods such as regression-based methods (e.g., multiple regression analysis, discriminatory analysis, logistic regression, variance analysis) and factor or cluster analysis are part of the key collection of statistical tools that can be used to either define or confirm theoretical hypothesis based on empirical data analysis. Many scientists in different fields have implemented one of these techniques to produce results that have influenced considerably the way we see the environment today, such as Spearman's (1904) job on overall psychological intelligence (factor analysis),Bedeian, Day, and Kelloway (1997) measurement error attenuation (SEM),Hofstede and Bond (1984) report on cross-cultural sociological variations (factor and cluster analysis) and (Altman, Iwanicz-Drozdowska, Laitinen, & Suvas, 2017) article on financial distress forecasting. A common impact for all of these techniques, however, is that they share three constraints, namely:
(a) The postulation of a straightforward model structure (at least for regression-based approaches);
(b) The hypothesis that all variables can be deemed observable; and
(c) The assumption that all variables are error-free measured, which may restrict their applicability in some situation studies.

With regard to the first hypothesis concerning the postulation of a straightforward model framework (i.e. one dependent and several autonomous variables), Jacoby (1978) indicated that "we live in a complicated, multivariate globe[ and] studying the effect of one or two factors in isolation would appear to be... relatively artificial and in consequential" (p. 91).

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

While model building always means omitting some aspect of reality (Shugan, 2002), this hypothesis of regression-based methods may be too limited for more complicated and realistic situations to be analyzed. This becomes particularly evident, for instance, when one wishes to explore the potential impact of *mediating or moderating* variables on the connection between one or more dependent and independent variables (for a comprehensive definition of these two terms, see (Baron & Kenny, 1986), which may result in some dependent variables affecting other dependent variables.. With regard to the second limitation, the assumption that all variables can be considered observable, McDonald (1996)emphasized that a variable can be called observable "if and only if its value can be obtained through an experiment in real-world sampling" (p. 239). Any variable that does not immediately correspond to anything that can be observed must therefore be regarded as non-observable (Dijkstra, 1983).

This definition makes it clear that only a handful of appropriate factors, such as age and gender, can be deemed observable, while "the impacts and characteristics of molecules, procedures, genes, viruses and bacteria are generally only noted indirectly" (S. Wold, 1993). With regard to the conjecture of factors measured without mistake, it should be borne in mind that each observation of the actual globe is accompanied by a certain measurement error, which may consist of two components (Bagozzi, Yi, & Phillips, 1991): (a) random mistake (e.g., triggered by the order of the questionnaire products or exhaustion of the respondent; Heeler and Ray (1972) and (b) systematic error, such as procedure variance Therefore, since the observed score of an item is always the sum of three parts, namely the true score of the variable, random error, and systematic error (Churchill Jr, 1979), first-generation techniques are strictly applicable only when there is no systematic or random error component, a rare situation in reality. To overcome these constraints of first-generation methods and other shortcomings, scientists have implemented partial Path Modeling Least Squares (PLS) to evaluate complicated interactions among latent variables. Many study fields have adopted the specific benefits of PLS route modeling, such as cognitive sciences e.g., (Bass & McKibben, 2003) as well as many company study fields such as marketing e.g.,(Hair, Sarstedt, Ringle, & Mena, 2012; Henseler, Ringle, & Sinkovics, 2009); strategy, Hulland (1999) organization (Sosik, Kahai, & Piovoso, 2009), and management data systems (Chin, Marcolin, & Newsted, 2003; Ringle, Sarstedt, & Straub, 2012) The popularity of PLS path modeling among researchers and professionals is due to four real benefits: First, PLS path modeling "does not involve any population or measurement scale assumptions" (Fornell & Bookstein, 1982). Thus, PLS route modeling can be used in extremely skewed distributions (Bagozzi & Yi, 1994), such as research on customer satisfaction (Fornell, 1995). Due to the rather smooth assumptions of PLS,H. Wold (1973), who created PLS route modeling, coined the word "soft modeling." Second, even with a tiny sample, it is possible to use PLS route modeling to assess interactions with multiple indices between latent variables (Chin & Newsted, 1999).

## Advantages of PLS Over other Methods of Analysis in Research

Since the PLS path modeling algorithm consists of ordinary least square regressions for distinct subpart of the focal path model, the general model's complexity hardly affects sample size demands. Third, contemporary, user-friendly PLS path modeling software with graphical user interfaces such as Smart PLS (Ringle et al. 2005), PLS-Graph (Soft Modeling Inc. 1992–2002) or the XLSTAT software PLS-PM module (Add in soft SARL 2007–2008) and open applications such as SEM-PLS (Monecke & Leisch, 2012) led to the attraction of PLS route modeling. Fourthly, PLS route modeling is preferred to covariance-based structural equation modeling (CBSEM) where inappropriate or non-convergent findings are probable (so-called instances of hey wood, c.f.(Krijnen, Dijkstra, & Gill, 1998; Reinartz, Haenlein, & Henseler, 2009), as in more complicated models where the amount of latent and visible variables is large in comparison to the number of observations and the number of indicators per latent variables is low. Furthermore, PLS is a smooth SEM modeling method with no information allocation assumptions (Vinzi, Chin, Henseler, & Wang, 2010). Thus, when the following situations are found, PLS-SEM becomes a useful solution to CB-SEM (Bacon, 1999; Hwang et al., 2010; Wong, 2010):

1. There is a tiny sample size.
2. Applications have little theory at their disposal.
3. Predictive precision is of paramount importance.
4. It is not possible to ensure correct model specification.

PLS Path Modeling is an assessment technique based on components (Tenenhaus, 2008). It is an iterative algorithm that solves the measuring model's blocks individually and then estimates the route coefficients in the structural model in a second step. Therefore, it is stated that PLS-PM explains at best the residual variance of the latent variables and, possibly, the manifest factors in any regression run in the model (Fornell & Bookstein, 1982). PLS Path Modeling is therefore regarded more as an exploratory method than as a confirmatory method. PLS-PM does not aim to reproduce the sample covariance matrix as opposed to the classical covariance-based strategy. PLS-PM is regarded a smooth modeling approach where there is no need for powerful assumptions (in terms of distributions, sample size and measurement scale). This is a very interesting characteristic particularly in those fields where, at least in complete, such assumptions are not tenable. On the other hand, this means a lack of the classic parametric inferential structure replaced by intervals of empirical confidence and hypothesis testing processes based on resampling methods (Chin, 1998; Tenenhaus, Vinzi, Chatelin, & Lauro, 2005) such as jackknife and bootstrap. It also results in less ambitious statistical characteristics for estimates, e.g. it is known that coefficients are biased but consistent overall (Cassel, Hackl, & Westlund, 1999). Lastly, PLS-PM is more geared towards optimizing projections (explained variances) than the estimates ' statistical precision. It is essential to note that PLS-SEM is not suitable for statistical analysis of all types. Marketers must also be conscious of certain PLS-SEM weaknesses, including:

1. If the sample size is tiny, highly valued structural path coefficients are required.
2. Multicollinearity problem unless treated well.
3. Because arrows are always single headed, undirected correlation cannot be modeled.

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

4. A potential absence of full consistency in results on latent variables can lead to partial estimation of components, loadings, and coefficients of trajectory.
5. In estimating the path coefficient loading, it can generate big mean square mistakes.

## Theoretical Concept of the Research

Scholars must transfer their suggested hypothesis to a statistical model in order to use PLS-PM (Rigdon, Sarstedt, & Ringle, 2017). This implies transferring the theoretical ideas and their hypothesized interactions into a structural model in the framework of SEM. "Theoretical concepts refer to ideas that share a certain unity or something. A theoretical definition describes the significance of a theoretical notion "(Bollen, 2012). Two kinds of theoretical concepts are distinguished: cognitive concepts, and design concepts, so-called artifacts. Theoretical ideas are typically depicted in the structural model by constructs (Rigdon, 2012). While constructs and latent variables are often equated (Bogozzi & Yi, 2012), we intentionally differentiate between a latent variable, i.e., a construct representing a notion of behavior, and an emerging variable, i.e., an artifact-representing construct. The operationalization of theoretical concepts, i.e. the specification of the theoretical concepts in the structural model, requires particular attention since estimates are likely to be inconsistent if the operationalization of a concept is not in line with the nature of the concept (Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016).

The aim of this article is to clarify the phases and metrics that scientists using PLS-SEM software can create model-based quality reporting and default outputs. When selecting to use PLS-SEM, we first summarize several original factors and cover elements such as sample sizes, reflective, formative distribution assumptions, reliability, validity, T-statistics and fitness testing. Then we address model assessment (exterior and internal models), including thumb laws, and introduce significant sophisticated alternatives that can be used. Next, we present several supplementary techniques for evaluating the robustness of the outcomes when it is to specification of the measurement model, nonlinear effects of the structural system, endogeneity and heterogeneity (J. F. Hair, Risher, Sarstedt, & Ringle, 2019; Latan, 2018). The multiple aspects that we discuss in the following parts are illustrated in Figure 1.

## Sample Size

It is certainly possible to use PLS-SEM with smaller samples, but the nature of the population determines the situations where small sample sizes are acceptable (Rigdon, 2016).The more heterogeneous the population, the bigger the sample size required to obtain an acceptable sampling error, if other situational features are equivalent(Cochran, 1977). If fundamental instructions on sampling theory are not regarded (Sarstedt et al., 2018), the findings are dubious. To determine the sample size required, researchers should rely on power analyzes that take into account the model structure, the expected meaning level, and the expected effect sizes (Marcoulides and Chin, 2013). Alternatively, Hair et al. (2017a) have recorded energy tables showing the sample sizes needed for a range of measurement and structural model features based on:

1. The level of importance
2. The strength of statistics
3. The minimum determination coefficient (R values) used in model
4. The maximum number of arrows pointing to a variable latent

In practice, a typical marketing research study would have a meaning level of 5 percent, 80 percent statistical power, and at least 0.25 R values. Using these parameters, you can look up the minimum sample size required from the guidelines suggested by Marcoulides & Saunders (2006), depending on the maximum number of arrows pointing to a latent variable as specified in the structural equation model (see Table 1):

**Table 1:** Suggested Sample Size in a Typical Marketing Research

| Minimum sample size required | Maximum # of arrows pointing at a latent variable in the model |
|---|---|
| 52 | 2 |
| 59 | 3 |
| 65 | 4 |
| 70 | 5 |
| 75 | 6 |
| 80 | 7 |
| 84 | 8 |
| 88 | 9 |
| 91 | 10 |

Sources (Hair *et al.,* 2013).

## Distributional Assumption

Many scientists say the lack of distributional assumptions is the primary reason to choose PLS-SEM (Nitzl, Roldan, & Cepeda, 2016); (do Valle and Assaker, 2016). While this is obviously a benefit in social science studies using PLS-SEM, which relies almost always on non-normal information alone, it is not a sufficient justification. Scholars have observed that estimating maximum likelihood with CB-SEM is robust against breaches of normality (Chou et al., 1991; Olsson et al., 2000), although much bigger sample sizes may be required (Boomsma and Hoogland, 2001). If the size of the data set is limited, CBSEM may produce abnormal results in non-normal data (Reinartz et al., 2009), whereas PLS-SEM shows higher robustness in such situations (Sarstedt et al., 2016b). It is remarkable that non-normal information may also influence PLS-SEM outcomes in a restricted amount of circumstances (Sarstedt et al., 2017a). For example, bootstrapping can generate peaked and skewed distributions with non-normal information. Using the bias-corrected and accelerated (BCa) bootstrapping routine handles this problem to some extent as it adjusts the skewed confidence intervals (Efron, 1987). Therefore, only selecting PLS-SEM for purposes of data distribution is not sufficient in most cases, but in conjunction with other reasons for using PLS-SEM it is definitely an benefit.

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

## Statistical Power

Researchers benefit from the elevated statistical power of the method compared to CB-SEM when using PLS-SEM (Reinartz et al., 2009; Hair et al., 2017b). Even when estimating common factor model information as presumed by CB-SEM, this feature holds (Sarstedt et al., 2016b). Greater statistical power implies that when they are actually present in the population, PLS-SEM is more likely to define interactions as significant (Sarstedt and Mooi, 2019). The greater statistical energy feature of PLS-SEM is quite helpful for exploratory studies examining less advanced or evolving theory. Wold (1985), defines the use of PLS-SEM as a "researcher-computer dialog." Tentative improvements to the model–such as the introduction of a fresh latent variable, an indicator or an internal relationship, or the omission of such an element–are evaluated for predictive significance [...] and the multiple pilot surveys are a quick and low-cost matter. "However, it is particularly important that PLS-SEM is not only suitable for exploratory research but also for confirmatory research (Hair et al., 2017a).

## Measurement Model Evaluation (Outer Models)

There are two types of measurement models in structural equation modeling; it can be formative or reflective.

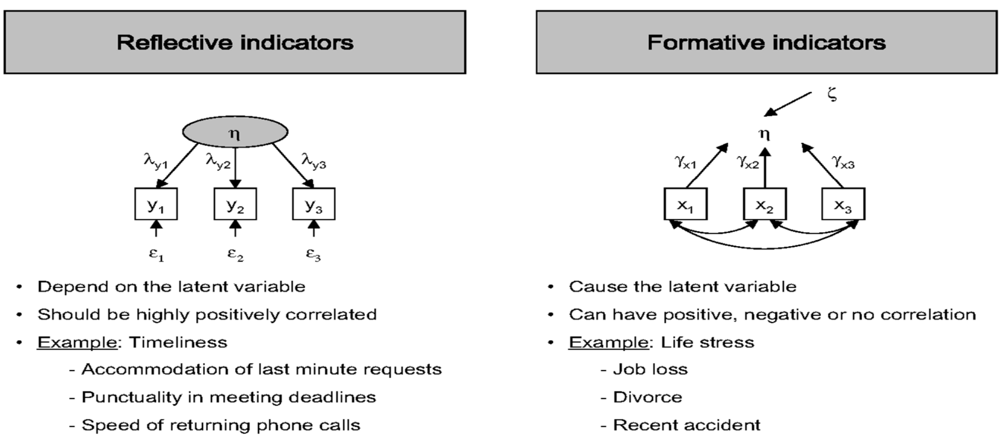


Figure 1: Reflective and formative indicators

## Reflective Measurement Models

The direction of causality ranges from the blue-color latent variable to the yellow-color indices in a reflective measurement scale. It is important to note that by default, Smart PLS assumes that when the model is built the indicators are reflective, with arrows pointing away from the latent blue-color variable (Figure 1). If the indices are extremely linked and interchangeable, they are reflective and should be carefully examined for their reliability and validity ((Haenlein & Kaplan, 2004; J. F. Hair, Ringle, & Sarstedt, 2013; Henseler & Sarstedt, 2013)

## Steps in Assessing Reflective Models

**Step #1:** Examines 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.

**Step #2**: Assessing the reliability of internal consistency, most often using the composite reliability of Jöreskog (1971). In general, greater values show greater reliability rates. For instance, reliability values between 0.60 and 0.70 are regarded "acceptable in exploratory research," values between 0.70 and 0.90 range from "satisfactory to good." Values of 0.95 and greater are difficult as they imply that the items are redundant, thus decreasing construct validity (Diamantopoulos et al., 2012; Drolet and Morrison, 2001). Reliability values of 0.95 and above also indicate the likelihood of unwanted reaction patterns (e.g. straight lining), causing inflated correlations between the error terms of the indicators. The alpha of Cronbach is another metric of quality of internal consistency that assumes comparable thresholds but generates reduced values than reliability of composites. In particular, the alpha of Cronbach is a less accurate measure of reliability because the items are unweighted. By comparison, with composite reliability, the weighted items are based on the individual loadings of the construct indicators and hence this reliability is greater than the alpha of Cronbach..

**Step #3**: The third stage of the evaluation of the reflective measurement model is the convergent validity of each measure of the construct. Convergent validity is the extent to which the construct converges in order to explain its items ' variance. The metric used to evaluate the convergent validity of a structure is the extracted average variance (AVE) for all elements on each structure. In order to calculate the AVE, the loading of each indicator on a construct must be squared and the mean value calculated. An acceptable AVE is 0.50 or greater, suggesting that at least 50% of the variance of its products is explained by the construct.

**Step #4:** To evaluate discriminating validity, the extent to which the structural model empirically distinguishes a structure from other constructs. Fornell and Larcker (1981) proposed the traditional metric and suggested that the AVE of each structure should be compared with that same construct's square inter-construction correlation (as a measure of shared variance) and all other reflectively measured constructs in the structural model. The shared variance should not be greater for all model constructs than their AVEs. However, recent study shows that this metric is not appropriate for discriminating validity evaluation. For example, Henseler et al. (2015) show that the Fornell Larcker criterion does not perform well, especially when there is only slightly different indicator loads on a construct (e.g. all indicator loads are between 0.65 and 0.85).

## Formative Measurement Models

Formative indicators occur in the model when it is necessary to reverse the direction of the arrows (see figure 1 above). That is, the arrow should point to the blue-color latent variable in SmartPLS from the yellow-color forming indicators. This can readily be achieved by

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

right-clicking the latent variable and choosing "Invert measurement model" to modify the direction of the arrow. Similarly, they are formative if the indices trigger the latent variable and are not interchangeable between themselves. In particular, there may be positive, negative, or even no correlations between these formative factors (Haenlein & Kaplan, 2004; Petter et al., 2007). As such, when using a formative measurement scale, there is no need to report indicator reliability, internal consistency reliability, and discriminating validity. This is because extracted exterior loads, composite reliability, and square root of average variance (AVE) are irrelevant for an uncorrelated measurement latent variable.

### Steps in Assessing Formative Models

Formative assessment models are assessed on the basis of: convergent validity, indicator collinearity, statistical significance, and object weight significance (Hair et al., 2017).

**Step #1**: Convergent validity is evaluated by the construct correlation with the alternative measure of the same notion. The method originally suggested by Chin (1998) is referred to as an assessment of redundancy. This includes the use as an exogenous latent variable of an current formative latent variable to predict an endogenous latent variable operationalized by one or more reflectively measured factors (see figure 2 below).

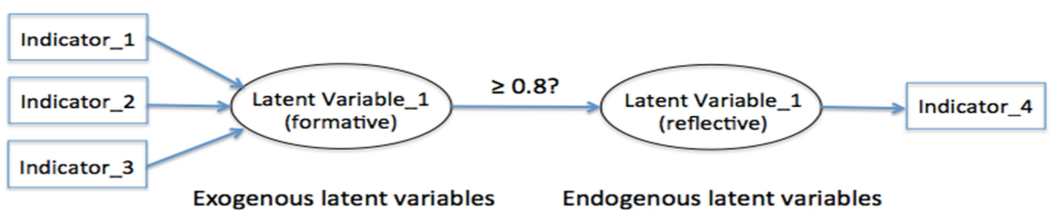


**Figure 2: Redundancy Analysis for Assessing Convergent Validity**

The reflective indicator ("Indicator_4" as in Figure 2) can be a global item in the `questionnaire that summarizes the essence of the latent variable the formative indicators ("Indicator_1", "Indicator_2", and "Indicator_3") intend to measure. For example, if the "Latent Variable_1" is about Corporate Social Responsibility, a survey question such as "Please evaluate to what degree this organization acted in a socially responsible way?" can be asked on a Likert scale of 0 (not a all) to 7 (completely), and this is the data for "Indicator_4".

**Step #2:** The variance inflation factor (IFV) is often used to assess the formation indicators ' collinearity. VIF values of 5 or higher show critical collinearity problems among the formally measured construct indices. Collinearity problems, however, may also happen at reduced VIF values of 3 (Mason and Perreault, 1991; Becker et al., 2015). The VIF values should ideally be nearly 3 and lower.

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

**Step #3:** To evaluate the significance and relevance (i.e. size). PLS-SEM is a non-parametric method and is therefore used to determine statistical significance by bootstrapping (Chin, 1998). In case the bootstrap distribution of the indicator weights is skewed, Hair et al. (2017) suggest using BCa bootstrap confidence intervals for meaning testing. Otherwise, the percentile technique should be used by scientists to build confidence intervals based on bootstrap (Aguirre-Urreta and Rönkkö, 2018). If an indicator weight confidence interval is zero, this shows that the weight is not statistically significant and that the indicator should be regarded for removal from the assessment model. According to Hair et al. (2017), if the loading is also not significant, indicators with a non-significant weight should definitely be eliminated. A small but significant loading of 0.50 and below indicates that if there is powerful support for its incorporation on the basis of measurement theory, one should consider removing the indicator.

**Step #4:** Researchers need to examine the significance of each indicator after evaluating the statistical significance of the indicator weights. The weights of the indicator are standardized to values between-1 and+ 1, but in rare instances values may also be smaller or greater than this, indicating an unusual outcome (e.g. due to collinearity problems and/or tiny sample sizes). A weight close to 0 shows a weak connection, while weights close to + 1 (or-1) show powerful beneficial (or negative) interactions.

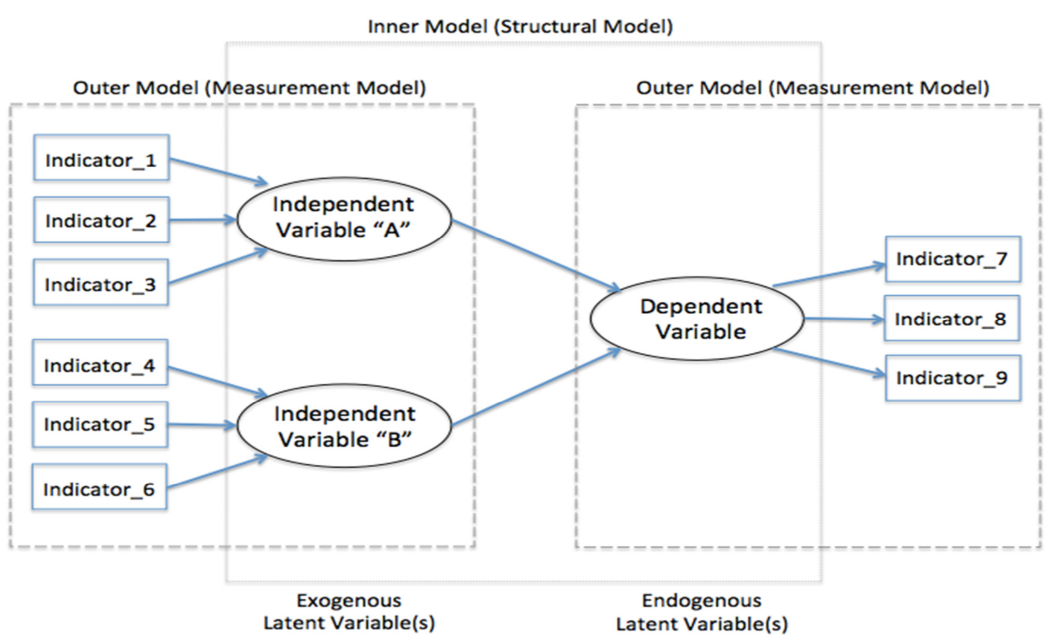


Figure 3: **Inner and Outer Model in a PLS-SEM Diagram**

## Structural Model Evaluation (Inner 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:

1. Overall fit of the estimated model,
2. The statistical significance and relevance of the path coefficients
3. the effect sizes (f2),
4. The coefficient of determination ($R^2$), and
5. The blindfolding-based cross validated redundancy measure $Q^2$

In addition, researchers should assess their model's out-of-sample predictive power by using the PLSpredict procedure (Shmueli et al., 2016). See Table 2 for more details on structural model assessment.

**Step #1: Evaluation of the Overall fit of the Estimated Model**

First, analysts should evaluate the overall fit of the estimated model through the bootstrap-based test of overall model fit and the SRMR as a measure of approximate fit to obtain empirical evidence for the proposed theory. Analysis in confirmatory research without assessing the overall model would be incomplete as this means ignoring empirical evidence for and also against the proposed model and the postulated theory. Without assessing the model fit, a researcher would not obtain any signal if he or she had incorrectly omitted an important effect in the model. Because the test for overall model fit was introduced only recently in the context of PLS-PM, the vast majority of models estimated by PLS-PM in past IS research has not been evaluated in this respect. However, because the overall model fit can now be tested in the context of PLS-PM, we encourage IS scholars to take this evaluation very seriously in causal research. In our example, all values of discrepancy measures were below the 95% quantile of their corresponding reference distribution ($HI_{95}$), indicating that the estimated model was not rejected at a 5% significance level. Moreover, the SRMR was below the preliminary suggested threshold of 0.080, indicating an acceptable model fit. This result suggests that the proposed model is well suited for confirming and explaining the development of social media capability and business process performance among firms. While the model fit suggests that there is a possibility that the world functions according to the specified model, the model can still be misspecified in the sense of over-parameterization, i.e., the model contains superfluous zero-paths Neither the bootstrap based test of model fit nor the SRMR punishes for unnecessary paths, i.e., neither of them rewards parsimony. Regardless of whether one conducts confirmatory or explanatory research, it remains indispensable to assess all path coefficients and their significance. Table 6 presents the construct correlation matrix.

**Step #2: Evaluation of Path Coefficients and their Significance Levels**

The path coefficient estimates are essentially standardized regression coefficients, whose sign and absolute size can be assessed. These coefficients are interpreted as the change in the dependent construct measured by standard deviations, if an independent construct is increased by one standard deviation while keeping all other explanatory constructs constant (ceteris paribus consideration).For example, increasing firm performance by one standard deviation will increase construction business growth by 0.623 standard deviations if all other variables are kept constant. Statistical tests and confidence

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

intervals can be used to draw conclusions about the population parameters. A path coefficient estimate is considered as statistically significant different from zero at a 5% significance level when its p-value is below 0.05 or when the 95% bootstrap percentile confidence interval constructed around the estimate does not cover the zero.

### Step #3: Evaluation of Effect Size ($f^2$)

The practical relevance of significant effects should be investigated by considering the effect sizes of the relationships between the constructs. The effect size is a measure of the magnitude of an effect that is independent of sample size. The $f^2$ values ranging from 0.020 to 0.150, 0.150 to 0.350, or larger or equal to 0.350, indicating weak, medium, or large effect size respectively. Just as all actors in a movie cannot play a leading role, it is unusual and unlikely that most constructs will have a large effect size in the model. We provide this clarification because scholars often expect/self-demand that all/most of their effect magnitude be large – an unrealistic expectation. This cautionary note extends to supervisors' expectations for their Ph.D. students.

### Step #4: Evaluation of Coefficient of Determination ($R^2$)

$R^2$ is used to assess goodness of fit in regression analysis. In the case of models estimated by OLS, the $R^2$ value gives the share of variance explained in a dependent construct. Thus, it provides insights into a model's in-sample predictive power. Moreover, R forms the basis for several innovative model selection criteria. Reporting $R^2$ makes PLS-PM research future-proof in this regard, because the new model selection criteria can still be calculated ex post as long as the $R^2$ values are given. The expected magnitude of $R^2$ depends on the phenomenon investigated. As some phenomena are already quite well understood, one would expect a relatively high $R^2$. For phenomena that are less well understood, a lower $R^2$ is acceptable. The $R^2$ values should be judged relative to studies that investigate the same dependent variable.

### Step #5 Evaluation of Redundancy Measure $Q^2$

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., 2017). 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.

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

Table 2: Summary of steps in evaluating structural model in PLS-SEM

| Steps | Description | Criterion | Suggested Threshold | Interpretation |
|---|---|---|---|---|
| **1. Overall fit of estimated mode** | Evaluating overall fit of the estimated model by evaluating discrepancy between the empirical indicator variance–covariance matrix and its model-implied counterpart | SRMR<br>$d_{ULS}$<br>$d_G$ | SRMR < 0.080 SRMR < $HI_{95}$<br>$d_{ULS} < HI_{95}$<br>$d_G < HI_{95}$ | Value of discrepancy measure below the 95% quantile of the corresponding reference distribution provides empirical evidence for the postulated model. In other words, it is possible that the empirical data stem from a world that functions as theorized by the model. |
| **2. Consider path coefficient estimates and their Significance levels** | Standardized regression coefficients are interpreted as change in standard deviations of the dependent variable if an independent variable is increased by one standard deviation while all other independent variables in the equation remain constant. | significance level Path coefficient estimates and their | Significant at 5% significance level, i.e., p-value <5% | Effect of independent variables on dependent variables is statistically significant |
| **3. Consider effect sizes ($f^2$)** | Measure of the magnitude of an effect that is independent of sample size. Give an indication about the practical relevance of an effect | $f^2$ value | $f^2 <$ 0.020: no substantial effect 0.020≤ $f^2$ < 0.15: weak effect size 0.15≤ $f^2$ < 0.350: medium effect size $f^2$ ≥ 0.350: large effect size | Degree of strength of an effect |

| | | | | |
|---|---|---|---|---|
| **4. Coefficient of determination ($R^2$)** | Explained variance of an dependent construct | $R^2$ | When the phenomena are already quite well understood, one would expect a high $R^2$. When the Phenomena are not yet well understood, a lower $R^2$ is acceptable. | Degree of variance explained for phenomenon under investigation |
| **5. Redundancy measure $Q^2$** | 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 | $Q^2$ | 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. | Indicating a higher predictive accuracy of the structural model. |

Source: Hair, Risher, Sarstedt and Ringle, (2018)

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

### A typical example on how to analyze and report data using SmartPLS software

The following customer satisfaction example will be used to demonstrate how to use the SmartPLS software application borrowed from Marketing Bulletin, 2013 for restaurant. Customer satisfaction is an example of a latent variable that is multidimensional and difficult to observe directly. However, one can measure it indirectly with a set of measurable indicators that serve as proxy. In order to understand customer satisfaction, a survey can be conducted to ask restaurant patrons about their dining experience. In this fictitious survey example, restaurant patrons are asked to rate their experience on a scale 10 representing four latent variables, namely Customer Expectation (EXPECT), Perceived Quality (QUAL), Customer Satisfaction (SAT), and Customer Loyalty (LOYAL), using a 7point Likert scales [(1) *strongly disagree*, (2) *disagree*, (3) *somewhat disagree*, (4) *neither agree nor disagree*, (5) *somewhat agree*, (6) *agree*, and (7) *strongly agree*]. The conceptual framework is visually shown in Figure 2, and the survey questions asked are presented in Table 3. Other than Customer Satisfaction (SAT) that is measured by one question, all other variables (QUAL, EXPECT, & LOYAL) are each measured by three questions. This design is in line with similar researches conducted for the retail industry (Hair et al., 2013).

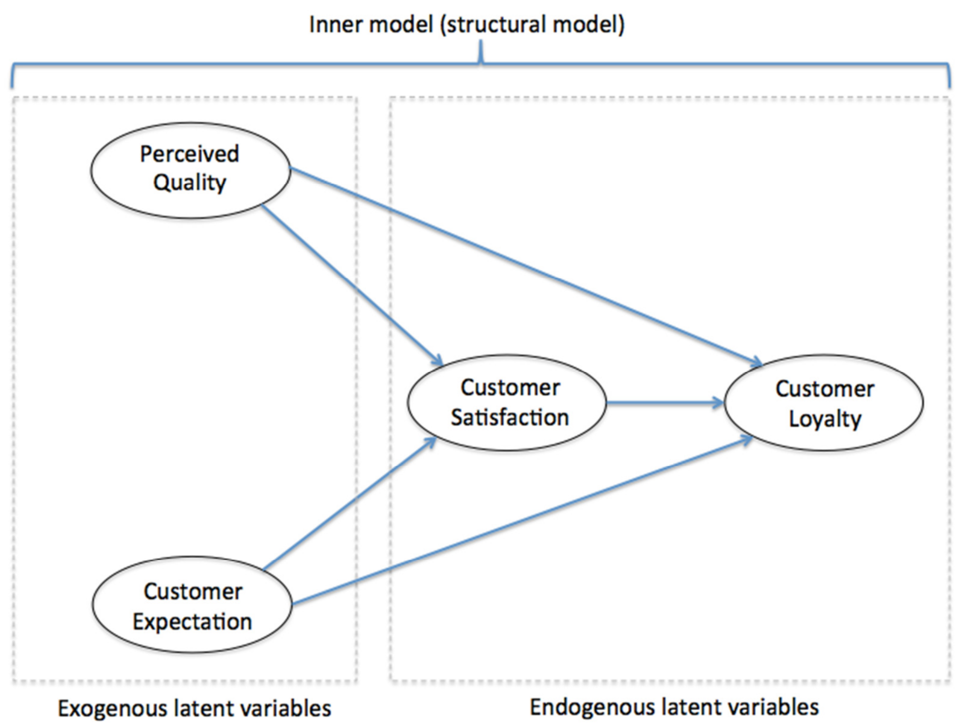


Figure 4: Conceptual Framework – Restaurant Example

Questions for Indicator Variables as in figure 4 above

Customer Expectation (EXPECT)
**expect_1** [this restaurant] has the best menu selection.
**expect_2** [this restaurant] has the great atmospheric elements.
**expect_3** [this restaurant] has good looking servers.

Perceived Quality (QUAL)
**qual_1** The food in [this restaurant] is amazing with great taste.
**qual_2** Servers in [this restaurant] are professional, responsive, and friendly.
**qual_3** [this restaurant] provides accurate bills to customers.

Customer Satisfaction (SAT)
**cxsat** If you consider your overall experiences with [this restaurant], how satisfied are you with [this restaurant]?

Customer Loyalty (LOYAL)
**loyal_1** I would recommend [this restaurant] to my friends and relatives.
**loyal_2** I would definitely dine at [this restaurant] again in the near future.
**loyal_3** If I had to choose again, I would choose [this restaurant] as the venue for this dining experience.

**Stage #1: Data Preparation for SmartPLS**
In this instance of the restaurant, the study information were typed manually into Microsoft Excel and saved as a format.xlsx (see Table 3). Without missing values, invalid observations or outliers, this information set has a sample size of 400. In order to ensure that Excel data can be properly imported by SmartPLS, the names of those indicators (e.g. expect 1, expect 2, expect 3) should be placed in the first row of an Excel spreadsheet and no "string" value (e.g., word or single point) should be used in other cells.

**Tale 3: Dataset from the Restaurant Example**

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | expect_1 | expect_2 | expect_3 | cxsat | loyal_1 | loyal_2 | loyal_3 | qual_1 | qual_2 | qual_3 |
| 2 | 2 | 6 | 5 | 6 | 2 | 6 | 7 | 5 | 4 | 2 |
| 3 | 3 | 5 | 4 | 5 | 3 | 5 | 5 | 2 | 1 | 2 |
| 4 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 5 | 4 | 4 | 5 | 6 | 5 | 6 | 6 | 5 | 2 | 3 |
| 6 | 5 | 7 | 6 | 7 | 7 | 7 | 7 | 6 | 6 | 3 |
| 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 8 | 7 | 5 | 7 | 7 | 7 | 7 | 7 | 4 | 1 | 7 |
| 9 | 6 | 6 | 6 | 4 | 5 | 4 | 6 | 4 | 3 | 4 |
| 10 | 5 | 7 | 6 | 6 | 5 | 7 | 7 | 7 | 5 | 7 |

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

Since SmartPLS cannot take native Excel file format directly, the data set has to be converted into *.csv* file format To do this, go to the "File" menu in Excel, and choose "CSV (Comma Delimited)" as the file format type to save it onto your computer (see Figure 5).

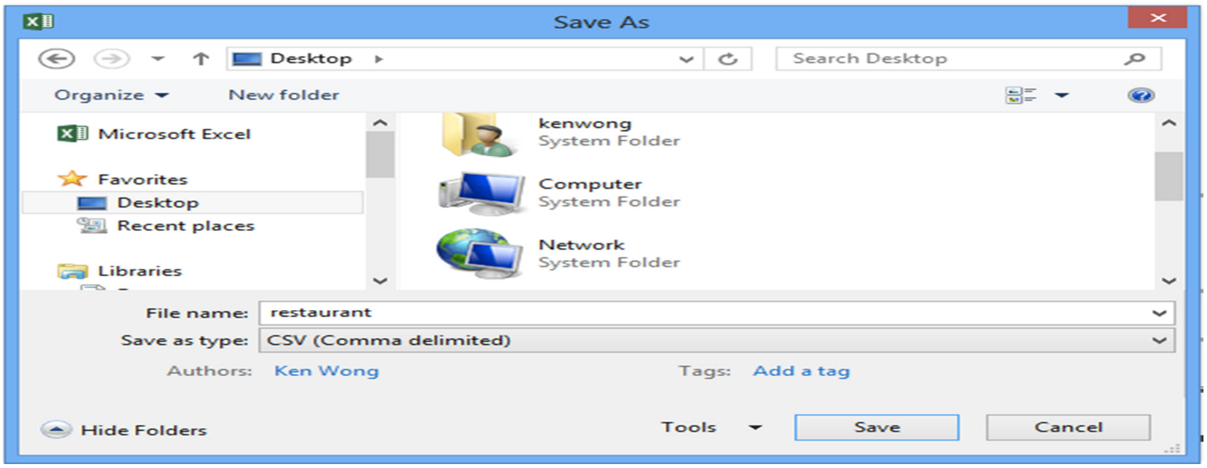


Figure 5: Save file as "CSV format" in Excel

**Stage #2: Project Creation in SmartPLS**

Now, start the program SmartPLS and go to the menu "File" to generate a fresh project. We'll name this project as a "restaurant" and then import the indicator data. Since there is no missing value in this restaurant data set, you can press the "Finish" button to create a PLS file. Once the data set is properly loaded into SmartPLS, click the "+" sign next to the restaurant to open the data in the "Projects" tab. (See Figure 6 above).

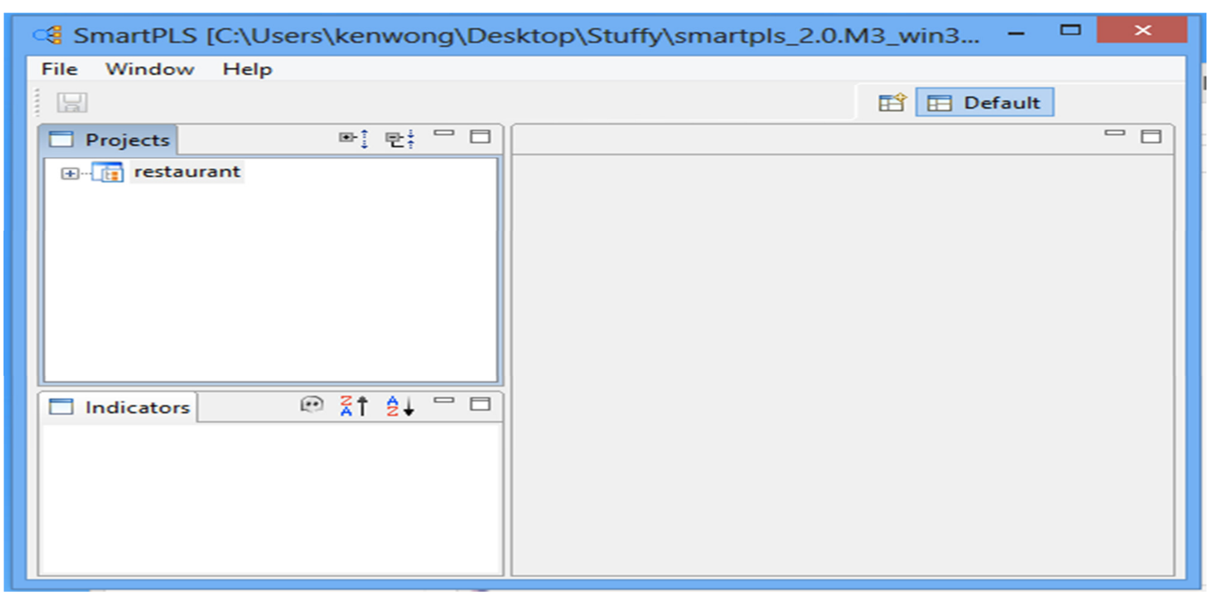


Figure 6: Project Selection window

Under the "restaurant" project directory, a "restaurant.splsm" PLS file and a corresponding "restaurant.csv" data file are displayed Click on the first one to view the manifest variables under the "Indicators" tab (see Figure 7).

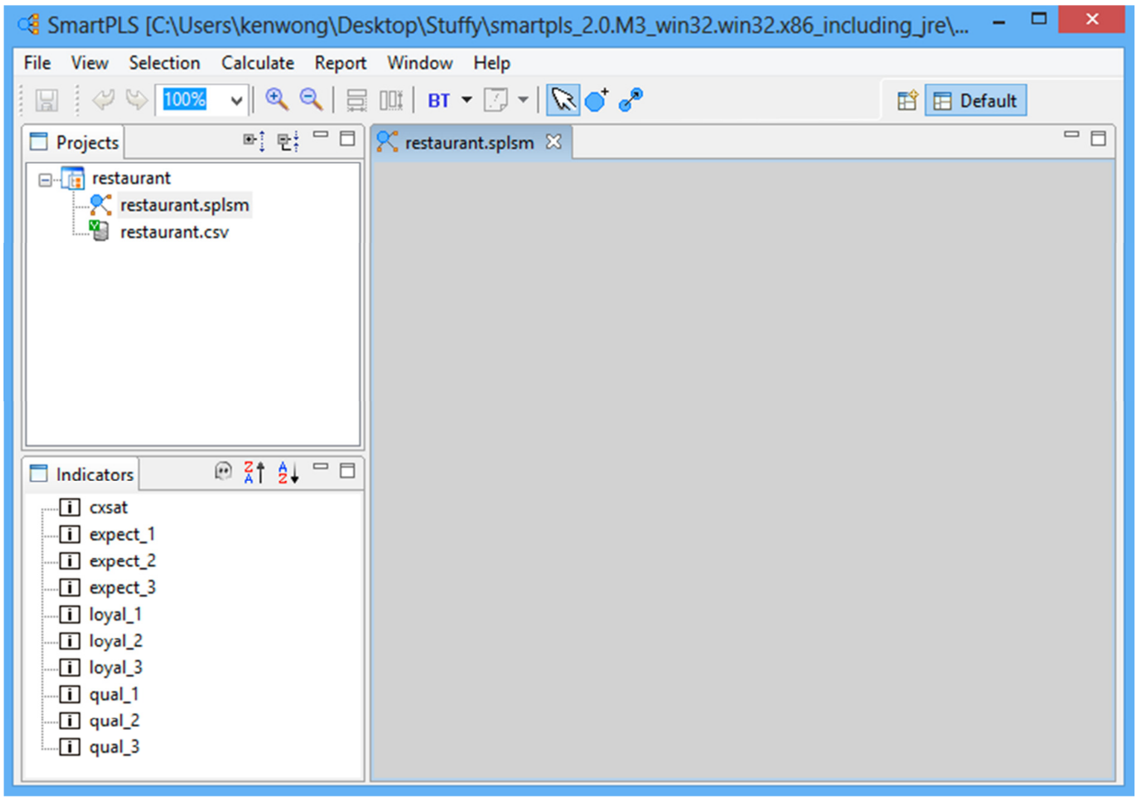

Figure 7: List of Indicators

### Stage #3: Building the Inner Model

Based on the conceptual framework as described in section 3 of this document, an inner model can be readily constructed in SmartPLS by first clicking on the modeling window on the correct side and then choosing the second last blue-colored circle icon entitled "Switch to Insert Mode." To generate the red-color circles representing your latent variables, click in the window. Once the circles are put, right-click on each latent variable to modify your model's default name into the variable name. To draw the arrows to link the variables together, press the last icon called "Switch to Connection Mode" (see Figure 8).

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

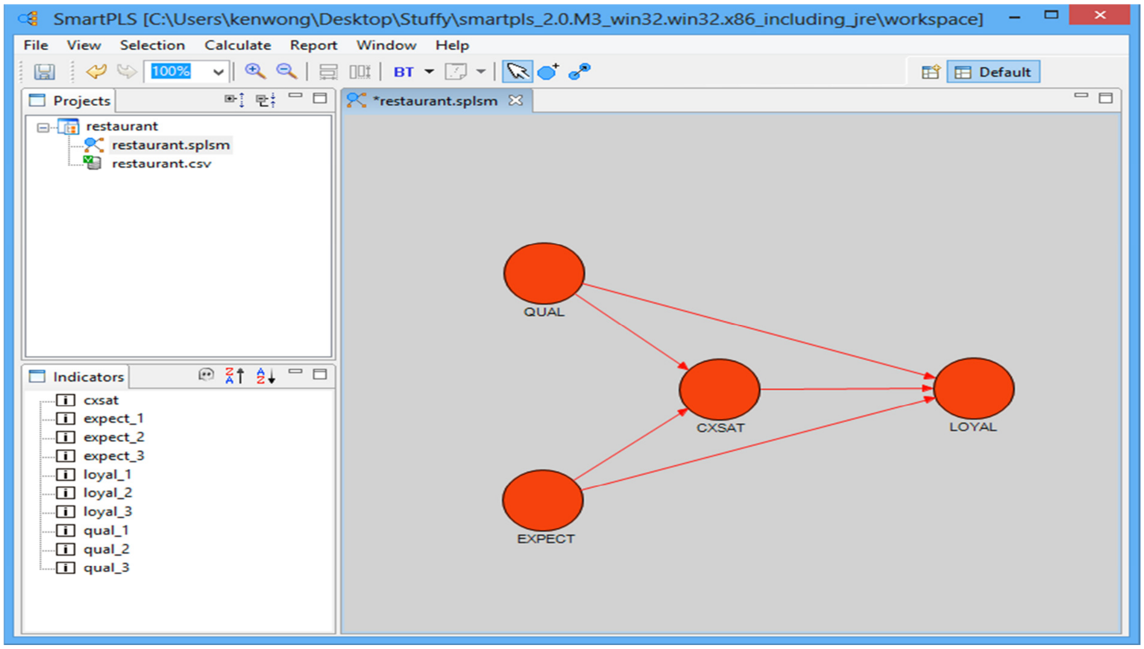


Figure 8: Building the Inner Model Based on Theoretical Concept

## Stage #4: Building the Outer Model

The next step is to construct the outer model. To do this, connect the variables to the latent variable by dragging them from the "Indicators" tab to the respective red circle one-by-one. Each indicator is depicted by a yellow rectangle and when the connection is formed, the color of the latent variable is shifted from red to blue. By using the "Align Top / Bottom / Left / Right" feature, the indicators can be readily moved to the screen, if you right-click the latent blue-color variable. It should look like the resulting model in Figure 9.

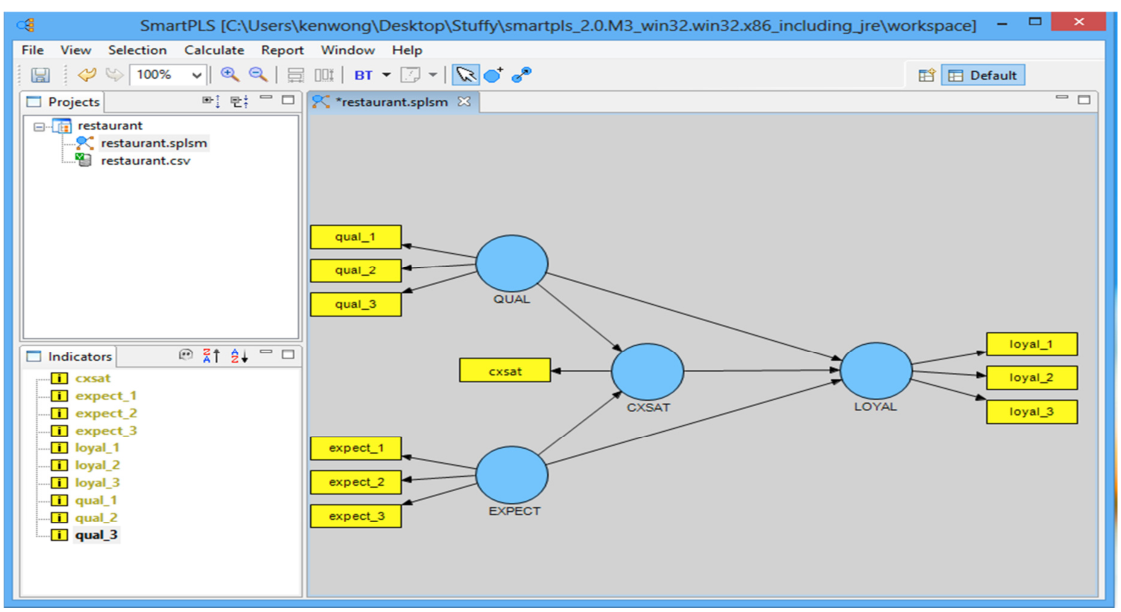


Figure 9: Building the Outer Model

**Stage #5: Running the Path-Modeling Estimation**

Once in SmartPLS the indicators and latent variables are effectively connected together (i.e. no more red-colored circles and arrows), the route modeling operation can be performed by going to the "Calculate" menu and choosing "PLS Algorithm." To activate it, simply press on the primary modeling window if the menu is dimmed. To display the default settings, a pop-up window will appear. Since our information set does not have a missing value, we continue straight to the bottom half of the pop-up window to configure the "PLS Algorithm Settings" with the following parameters (see Figure 10):

1. Weighting Scheme: Path Weighting Scheme
2. Data Metric: Mean 0, Variance 1
3. Maximum Iterations: 300
4. Abort Criterion: 1.0E-5
5. Initial Weights: 1.0

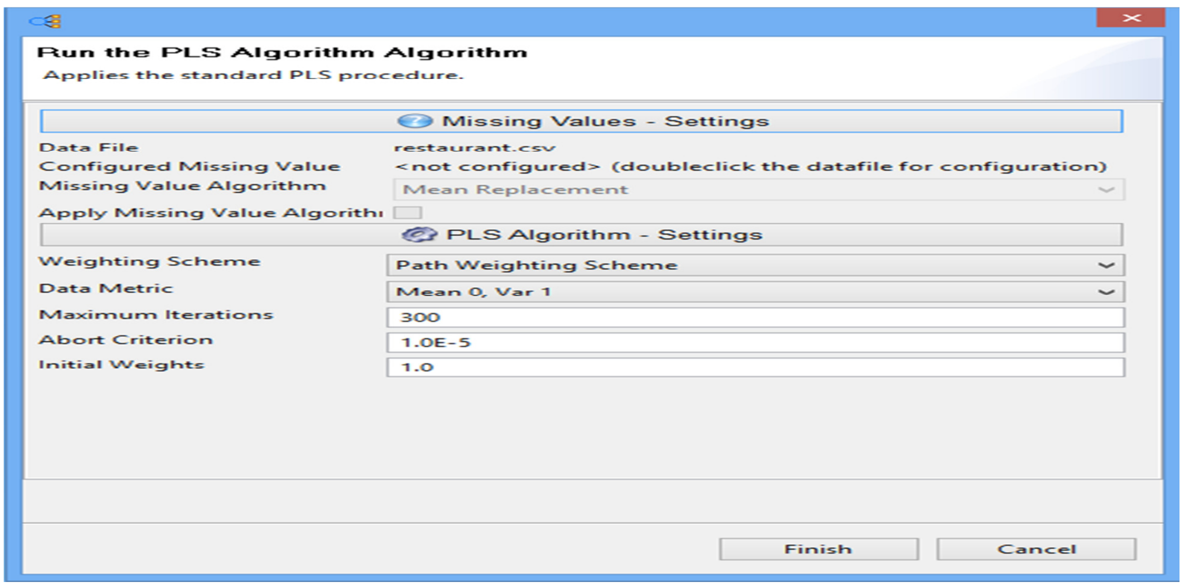


Figure 10: Configuring the PLS Algorithm

To run the path modeling, press the "Finish" button. There should be no error messages popping up on the screen, and the result can now be assessed and reported.

**Stage #6: Assessing and Reporting the PLS-SEM Output**

Some fundamental components should be discussed in your study report for an original evaluation of the PLS-SEM model. Stage 1-4 in chapter 4 should be discussed in detail for a reflective measurement model. As an instance, the outcome of the restaurant was provided in Figure 11, showing all variables with their track coefficient for both indices and latent variables.

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

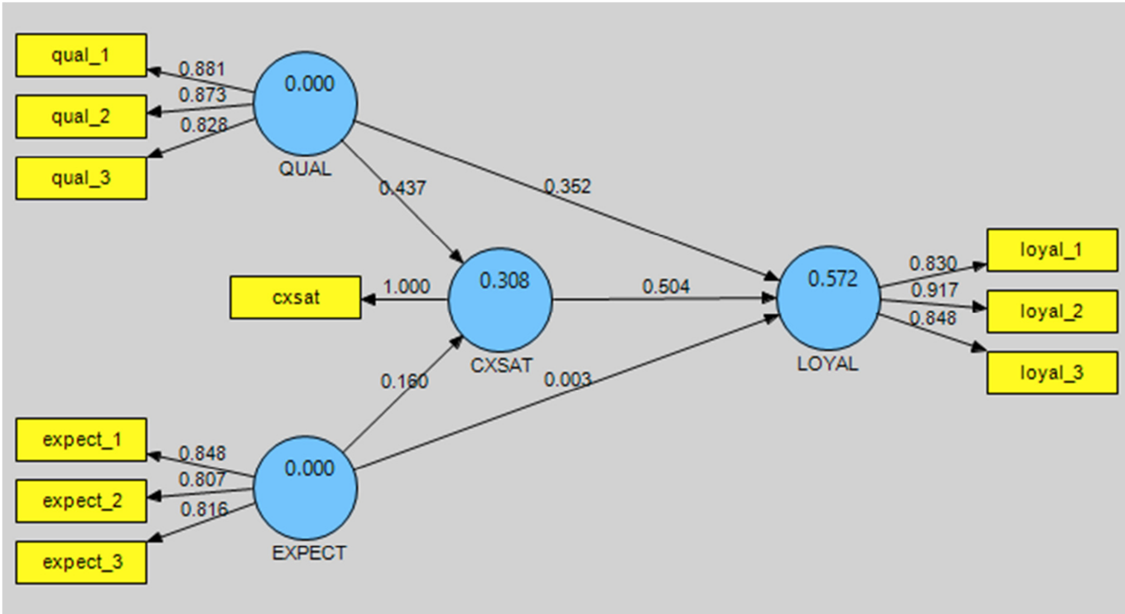


Figure 11: PLS-SEM Output Result

SmartPLS presents path modeling estimations not only in the Modeling Window but also in a text-based report which is accessible via the "Report" menu. In the PLS-SEM diagram, there are two types of numbers:

1. Numbers in the circle: These show how much the variance of the latent variable is being explained by the other latent variables.
2. Numbers on the arrow: These are called the path coefficients. They explain how strong the effect of one variable is on another variable. The weight of different path coefficients enables us to rank their relative statistical importance.

## Measurement Model (Outer Model)

Table 5: Summary of the Reflective Outer model result

| Latent | Indicators | Loadings Indicator | Indicator Reliability (i.e Loadindg$^2$) | Composite Reliability (CR) | AVE |
|---|---|---|---|---|---|
| QUAL | Qual_1 | 0.881 | 0.776 | 0.8958 | 0.7415 |
| | Qual_ | 0.873 | 0.762 | | |
| | Qual_3 | 0.828 | 0.686 | | |
| EXPECT | Expect_1 | 0.848 | 0.719 | 0.8634 | 0.6783 |
| | Expect_2 | 0.807 | 0.651 | | |
| | Expect_3 | 0.816 | 0.666 | | |
| LOYAL | Loyal_1 | 0.830 | 0.689 | 0.8995 | 0.7494 |
| | Loyal_2 | 0.917 | 0.841 | | |
| | Loyal_3 | 0.848 | 0.719 | | |

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

**Indicator loadings**: Observed variables with an outer load of 0.7 or greater are considered to be highly acceptable while the outer load should be discarded with a value of less than 0.7. Nevertheless, the cut-off value accepted for the outer loading was 0.708 for this research. The outer loads ranged from 0.807 to 0.917 from Table 5. This shows that more than 50 percent of the variance of the indicator is explained by the structure, thus giving acceptable item reliability. Similarly, the reliability of the indicator is greater than the minimum acceptable range of of 0.4 and near the preferred rate of 0.70 for three indices, whereas it is higher than 0.7 for five indices as shown in Table 5.

**Internal consistency reliability**: Table 5 demonstrates that for all constructs the composite reliability (CR) was higher than 0.80. The CR showed that the scales were reasonably reliable and indicated that the minimum threshold level of 0.70 was exceeded by all latent construct values. This demonstrates that the indicators are appropriate for restaurant exploratory research. Research has eliminated the limitations of the unwanted pattern of response and the inflated correlation between the terms of mistake.

**Convergent Validity (AVE):** The Average Variance Extracted (AVE) of each latent construct was calculated by squaring the mean of indicators for that construct to confirm the convergent validity of the factors. Table 5 shows that all AVE values were more than 0.5, so for this research model convergent validity was confirmed. The findings confirmed that the structure explains at least 50% of the variance of its goods.

**Discriminant Validity**: The next effort was the latent constructs ' discriminating validity. Discriminant validity defines that in any construct the manifest variable is distinct from other constructs in the path model, where its cross-load value in the latent variable is greater than in any other construct [72]. The criterion and cross-loadings of Fornell and Larcker (1989) were used to assess the discriminating validity[70]. The standard proposed is that a construct should not display the same variance as any other construct that exceeds its AVE value [72]. Table 6 shows the model's Fornell and Larcker criterion test where the squared correlations were compared to the correlations of other latent buildings. Table 6 demonstrates that all correlations were lower relative to the average square root variance exerted along the diagonals, suggesting adequate discriminating validity. This showed that in each structure the observed factors indicated the specified latent variable confirming the model's discriminating validity.

**Table 6: Fornell-Larcker Criterion Analysis for Checking Discriminant Validity**

| | QUAL | EXPECT | CXSAT | LOYAL |
|---|---|---|---|---|
| QUAL | **0.861** | | | |
| EXPECT | 0.655 | **0.824** | | |
| CXSAT | 0.542 | 0.446 | Single Item | |
| LOYAL | 0.626 | 0.458 | 0.695 | **0.866** |

As a result, the suggested conceptual model was supposed to be acceptable, with confirmation of adequate reliability, convergent validity, and discriminant validity and the verification of the research model.

**Structural Model (Inner Model)**
We confirmed the validity and reliability of the measurement model. The next stage was to assess the results of the internal structural model. This included observing the predictive relevance of the model and the relationships between the constructs. The determination coefficient ($R^2$), the path coefficient (β -value) and the T-statistic value, the effect size ($f^2$), the model's predictive significance ($Q^2$), and the goodness-of-fit (GOF) index are the main norms for the internal structural model evaluation. Section 5 (step 1-5) above described the detailed debate of each item.

**Coefficient of determination ($R^2$):** The coefficient of determination measures the overall effect size and variance explained in the endogenous construct for the structural model and is thus a measure of the model's predictive accuracy. In current study, the inner path model was 0.572 (figure 11) for the loyalty explain 57.20% of the variance in the loyalty, meaning that about 57.20% of the change in the costumer loyalty was due to three latent constructs in the model (Quality, Expectation and Satisfaction). According to Henseler et al. (2017), and Hair et al., (2016) an $R^2$ value of 0.572 is considered substantial, an $R^2$ value of 0.50 is regarded as moderate, and an $R^2$ value in this study was substantial. Similarly, quality and expectation together explain 30.8% of variance in customer satisfaction (CXSAT), which indicates as a weak determination.

**Measuring the Effect Size($f^2$):** The $f^2$ is the degree of the impact of each exogenous latent construct on the endogenous latent construct. When an independent construct is deleted from the path model, it changes the value of the coefficient of determination ($R^2$) and defines whether the removed latent exogenous construct has a significant influence on the value of the latent endogenous construct. The $f$ values were 0.35 (strong effect), 0.15 (moderate effect), and 0.02 (weak effect). The effect size for quality, expectation and customer satisfaction on loyalty were 0.201, 0.268, and 0.257, respectively. Hence, the values were (moderate effect),. According to Cohen's recommendation, the $f$ of all three exogenous latent constructs on customer loyalty had a moderate effect on the value of $R^2$. Furthermore, all the three independent latent constructs in this study participated relatively to the greater $R^2$ value (57.2%) in the dependent variable.

**Path Coefficient β- Value and T-Statistic Value**: The path coefficients in the PLS and the standardized β coefficient in the regression analysis were similar. Through the β value, the significance of the hypothesis was tested. The β denoted the expected variation in the dependent construct for a unit variation in the independent construct(s). The β values of every path in the hypothesized model were computed, the greater the β values, the more the substantial effect on the endogenous latent construct. However, the β value had to be verified for its significance level through the T-statistics test. The bootstrapping procedure was used to evaluate the significance of the hypothesis. To test the significance of the path coefficient and T-statistics values, a bootstrapping procedure using 5000 subsamples with no sign changes was carried out for this study as explained below:

Go to the "Calculate" menu and select "Bootstrapping". In SmartPLS, sample size is known as Cases within the Bootstrapping context, whereas the number of bootstrap subsamples is known as Samples. Since there are 400 valid observations in our restaurant data set, the number of "Cases" (not "Samples") in the setting should be increased to 400 as shown in Figure 12. The other parameters remain unchanged:
1. Sign Change: No Sign Changes
2. Cases: 400
3. Samples: 5000

It worth noting that if the bootstrapping result turns out to be insignificant using the "No Sign Changes" option, but opposite result is achieved using the "Individual Sign Changes" option, you should subsequently re-run the procedure using the middle "Construct Level Changes" option and use that result instead. This is because this option is known to be a good compromise between the two extreme sign change settings.

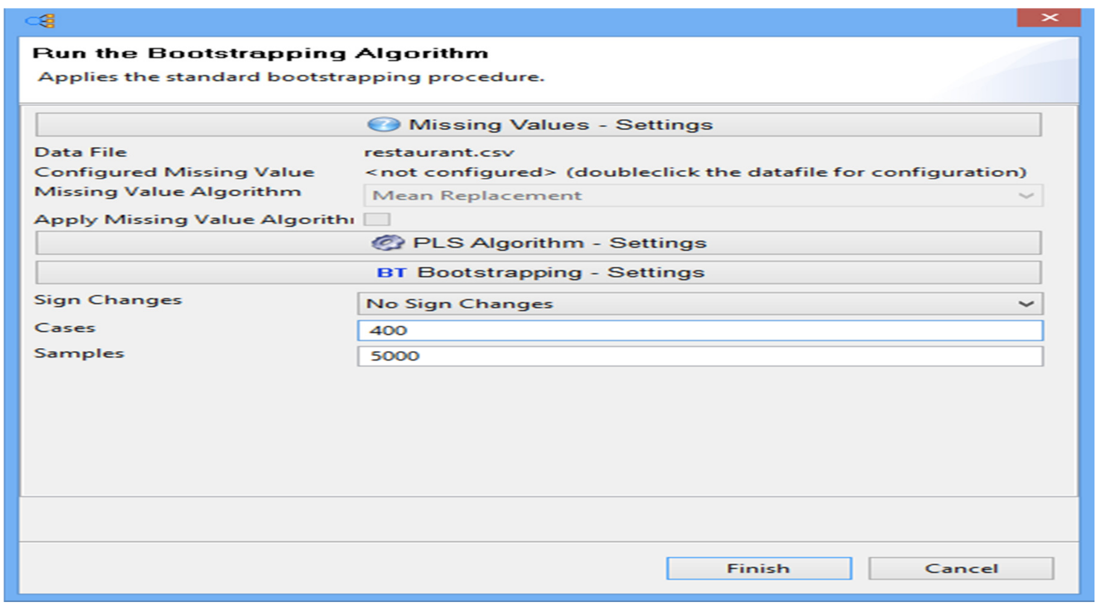


Figure 12: Bootstrapping Algorithm

Go to the "Path Coefficients (Mean, STDEV, T-Values) window within the Default Report bootstrapping section once the bootstrapping procedure is completed. Check the figures in the column "T-Statistics" to see whether or not the inner model's route coefficients are important. Using a two-tailed t-test with a meaning point of 5 percent, if the T-statistics are greater than 1.96, the path coefficient will be important. It can be seen in our restaurant instance that only the connection "EXPECT–LOYAL" (0.0481) is not important. This proves our previous findings visually when viewing the outcomes of PLS-

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

SEM (see Figure 11). All other internal model route coefficients are statistically important (see Figure 13 and Table 7).
All other path coefficients in the inner model are statistically significant (see Figure 13 and Table 7)

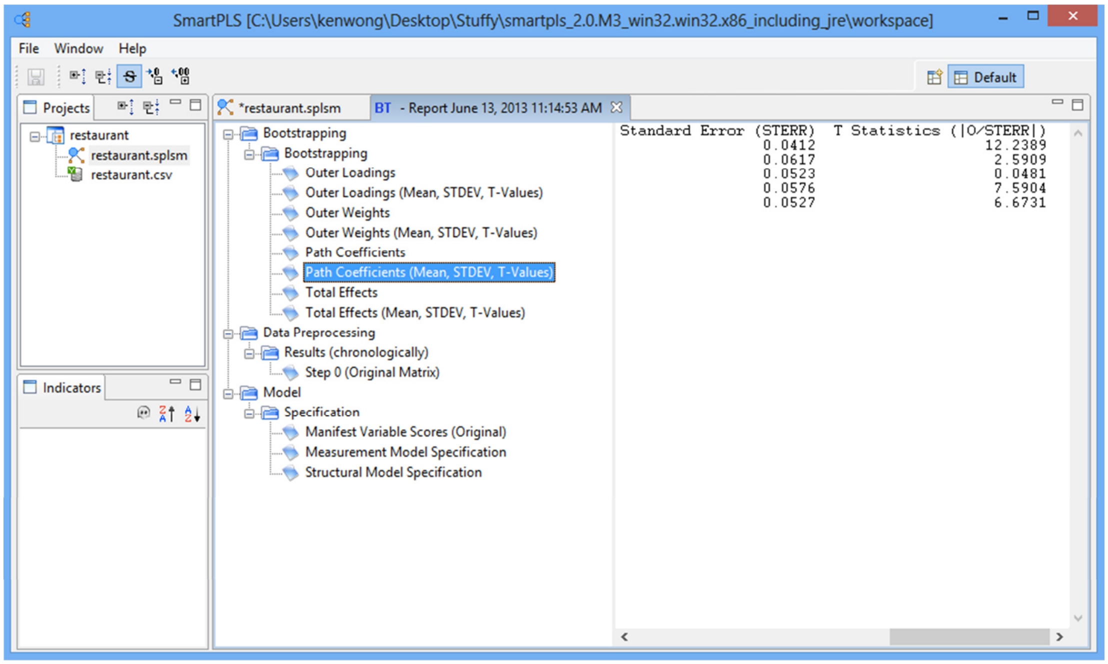


Figure 13: Bootstrapping Results - Path Coefficients for Inner Model

Table 7: T-Statistics of Path Coefficients (Inner Model)

| Path | □□□□□□□□□□□□□□□□□□□□□□□□□ | T-Statistic | p-Value |
|---|---|---|---|
| CXSAT □ LOYAL | 0.504 | 12.2389 | 0.00 |
| EXPECT □ CXSAT | 0.160 | 2.5909 | 0.00 |
| EXPECT □ LOYAL | 0.003 | 0.0481 | 0.72 |
| QUAL □ CXSAT | 0.437 | 7.5904 | 0.03 |
| QUAL □ LOYAL | 0.352 | 6.6731 | 0.021 |

In $H_1$, we anticipated that customer satisfaction would influence customer loyalty significantly and positively. The findings in Table7and Figure 11 confirmed that client satisfaction had a significant impact on client loyalty (β= 0.504, T= 12.2389, p < 0.000). $H_1$ has therefore been robustly endorsed. In addition, when observing the direct and positive influence of client expectations on client loyalty ($H_2$), the findings from Table7and Figure11 confirmed that client expectations strongly influenced client loyalty (β= 0.160, T= 2.5909, p < 0.000) and $H_2$. The influence of client expectations on client loyalty was beneficial and non-significant (β= 0.003, T= 0.0481, T= 0.0481). The impact of food quality on client loyalty (β= 0.437, T= 7.5904, p < 0.000) was significant, thus promoting $H_4$. Similarly, the findings in Table 7 given empirical assistance for $H_5$, where the influence of quality food on client loyalty was positive and the client loyalty was significantly impacted (β= 0.352, T= 6.6731, p < 0.021), confirming hypothesis ($H_5$). The higher the beta coefficient, the higher the impact of a latent exogenous structure on the latent endogenous structure. Table7and Figure11 showed that, compared to other β values in the model, the customer satisfaction had the highest route coefficient of β= 0.504, which indicated that it had a higher variance value and a strong impact on the improvement of restaurant services. The customer expectation had the least impact on the performance of the project with β= 0.003.

## DISCUSSIONS AND IMPLICATIONS OF FINDINGS

The aim of this instance is to show how restaurant managers can enhance their company by understanding client expectation (EXPECT), perceived quality (QUAL), client satisfaction (SAT) and client loyalty (LOYAL) relationships. The significant factors leading to customer loyalty are recognized through a study of the restaurant staff and the subsequent structural equation modeling in SmartPLS. Customers are discovered to care about the taste of food, table service, and precision of bills in this studies. They are excellent indicators of perceived quality (QUAL) with loadings of 0.881, 0.873 and 0.828 respectively. Restaurant management should not ignore these fundamental aspects of daily operation as perceived quality has been shown to have a significant impact on the level of satisfaction of clients, their intention to return, and whether or not they would suggest this restaurant to others.

In the meantime, it is also disclosed that menu choice, atmospheric components and good-looking employees are significant client expectation indicators (EXPECT), with loads of 0.848, 0.807 and 0.816 respectively. While meeting these client expectations can keep them satisfied, improvement in these fields does not have a significant impact on customer loyalty due to its weak effect (0.03) in the connection. Management should therefore only allocate funds to enhance these regions after care has been taken of food taste, table service and bill precision. The internal model assessment demonstrates that perceived quality (QUAL) and client expectations (EXPECT) together can only explain the variance of customer satisfaction (CXSAT) by 30.8 percent. It is an significant finding because it indicates that other factors should be considered in future studies by restaurant executives when exploring customer satisfaction.

### Other Considerations When Conducting an In-depth Analysis of PLS-SEM

The depth of the analysis of PLS-SEM depends on the scope of the research project, the complexity of the model and the common presentation in the preceding literature. For instance, a thorough evaluation of PLS-SEM would often include an evaluation of multicollinearity. That is, in the inner model, each set of exogenous latent variables is checked for potential collinearity problem to see if any variables should be eliminated, merged into one, or simply developed a latent higher-order variable. The latent variable scores (PLS→ Calculation Results→Latent Variable Scores) can be used as input for multiple regression in IBM SPSS Statistics to obtain the tolerance or Variance Inflation

International Journal of Environmental Studies and Safety Research
ISSN: 2536-7277 (Print): 2536-7285 (Online)
Volume 4, Number 4, December 2019
http://www.casirmediapublishing.com

Factor (VIF) values, as these figures are not provided by SmartPLS. First, make sure that the information set is in the format of the.csv file. Then, import the information into SPSS and go to Linear Analysis-Regression. The exogenous latent variables (the predictors) are arranged as independent variables in the SPSS linear regression module, while the dependent variable is configured as another latent variable (which does not behave as a predictor). VIF is a "1/Tolerance" calculation. In general, we need a VIF of 5 or lower (i.e., 0.2 or higher tolerance level) to prevent the issue of collinearity (Hair et al., 2011). In addition to checking collinearity, the f-effect size of the model can be discussed in detail, showing how much an exogenous latent variable contributes to the R value of an endogenous latent variable. Effect size evaluates the magnitude or strength of the connection between the latent variables in easy terms. Such debate may be crucial as effect size helps scientists evaluate a research study's general contribution. Chin, Marcolin, and Newsted (1996) made it clear that researchers should not only specify whether or not the connection between factors is important, but also report the magnitude of the effects between factors. Meanwhile, another element that can be studied for the inner model is predictive significance.

The values of Stone-Geisser (Q2) (i.e. cross-validated redundancy measures) can be acquired in SmartPLS (Calculate→ Bindfolding) through the Bindfolding operation. For most studies, an omission distance (OD) of 5 to 10 is suggested in the Bindfolding setting window (Hair et al., 2012). It is also possible to calculate and discuss the q2 effect size for the Q values. If there is a mediating latent variable in the model, the total effect of a specific exogenous latent variable on the endogenous latent variable can also be discussed. Total Effect value can be discovered in the default document (PLS→ Criteria for Quality→Total Effects). In the Bootstrapping procedure (Bootstrapping-→Total Effects (Mean, STDEV, T-Values)) the significance of Total Effect can be tested using T-Statistics. In addition, unobserved heterogeneity may need to be evaluated when there is little information about the underlying data as it may impact the validity of the assessment of PLS-SEM.

## CONCLUSION

PLS-SEM is increasingly being used to predict structural equation models (Hair et al., 2014). Scholars need a detailed and concise overview of the factors and indicators needed to ensure that their evaluation and reporting of PLS-SEM results are accurate–before submitting their article for review. Such reporting guidelines have been given for prior research (Hair et al., 2011; Hair et al., 2013; Hair et al., 2012b; Chin, 2010; Tenenhaus et al., 2005; Henseler et al., 2009) that, in the light of more recent research and methodological advances in the field of PLS-SEM, need to be continually expanded and modified. We hope this paper will achieve this goal. To researchers who have not used PLS-SEM in the past, this article is a good point of orientation when writing and finalizing their manuscripts. In addition, this is a good over view and reminder of how to prepare PLS-SEM manuscripts for researchers experienced in the application of PLS-SEM. This knowledge is also essential for reviewers and journal editors to ensure the rigor of the PLS-SEM studies published. We provide a summary of a variety of recently proposed enhancements (PLS predict and Model Comparison Metrics) as well as additional robustness control methods (e.g. measurement and structural models evaluation), that we suggest should be used–if necessary–when using PLS-SEM. Eventually, while a few researchers have published negative articles on the use of PLS-SEM, more recently a number of prominent researchers have recognized the value of PLS as a SEM technique (Petter, 2018). We believe that social science scholars would be dissatisfied if they did not use all the 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 PLS using bootstrap confidence intervals *MIS Quarterly, 42 (3), 1001-1020*

Altman, E. I., Iwanicz‐Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z‐score model. *Journal of International Financial Management & Accounting, 28*(2), 131-171.

Bagozzi, R. P., & Yi, Y. (1994). Advanced topics in structural equation models. *Advanced methods of marketing research, 151*.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative science quarterly*, 421-458.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology, 51*(6), 1173.

Bass, A. H., & McKibben, J. R. (2003). Neural mechanisms and behaviors for acoustic communication in teleost fish. *Progress in neurobiology, 69*(1), 1-26.

Becker, J. M., Ringle, C. M., Sarstedt, M., & Völckner, F. (2015). How collinearity affects mixture regression results. *Marketing Letters, 26*(4), 643-659.

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement, 57*(5), 785-799.

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. *Structural equation models: Present and future. A Festschrift in honor of Karl Jöreskog, 2*(3), 139-168.

Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology, 38*, 37-72.

Cassel, C., Hackl, P., & Westlund, A. H. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of applied statistics, 26*(4), 435-446.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research, 295*(2), 295-336.

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information systems research, 14*(2), 189-217.

Chou, C. P., Bentler, P. M., & 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.

Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. *Statistical strategies for small sample research, 1*(1), 307-341.

Churchill Jr, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of marketing research, 16*(1), 64-73.

Cochran, W. G. (1977). Sampling techniques-3.

Diamantopoulos, E., & Durner, W. (2012). Dynamic nonequilibrium of water flow in porous media: A review. *Vadose Zone Journal, 11*(3).

Do Valle, P. O., & 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, 55*(6), 695-708.

Drolet, A. L., & Morrison, D. G. (2001). Do we really need multiple-item measures in service research?. *Journal of service research, 3*(3), 196-204.

Dijkstra, T. (1983). Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics, 22*(1-2), 67-90.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American statistical Association, 82*(397), 171-185.

Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika, 61*(1), 101-107.

Fornell, C. (1995). The quality of economic output: Empirical generalizations about its distribution and relationship to market share. *Marketing science, 14*(3_supplement), G203-G211.

Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of marketing research, 19*(4), 440-452.

Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding statistics, 3*(4), 283-297.

Hair, Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). 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.

Hair, J. F., Ringle, C. M., & 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, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2-24.

Heeler, R. M., & Ray, M. L. (1972). Measure validation in marketing. *Journal of marketing research, 9*(4), 361-370.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing *New challenges to international marketing* (pp. 277-319): Emerald Group Publishing Limited.

Henseler, J., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics, 28*(2), 565-580.

Hofstede, G., & Bond, M. H. (1984). Hofstede's culture dimensions: An independent validation using Rokeach's value survey. *Journal of cross-cultural psychology, 15*(4), 417-433.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic management journal, 20*(2), 195-204.

Jacoby, J. (1978). Consumer Research: How valid and useful are all our consumer behavior research findings? A State of the Art Review1. *Journal of marketing, 42*(2), 87-96.

Krijnen, W. P., Dijkstra, T. K., & Gill, R. D. (1998). Conditions for factor (in) determinacy in factor analysis. *Psychometrika, 63*(4), 359-367.

Latan, H. (2018). PLS path modeling in hospitality and tourism research: the golden age and days of future past *Applying Partial Least Squares in Tourism and Hospitality Research* (pp. 53-83): Emerald Publishing Limited.

Marcoulides, G. A., & Chin, W. W. (2013). You write, but others read: Common methodological misunderstandings in PLS and related methods. In *New perspectives in partial least squares and related methods* (pp. 31-64). Springer, New York, NY.

Marcoulides, G. A., & Saunders, C. (2006). Editor's comments: PLS: a silver bullet?. *MIS quarterly*, iii-ix.

Mason, C. H., & Perreault Jr, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of marketing research, 28*(3), 268-280.

McDonald, R. P. (1996). Path analysis with composite variables. *Multivariate Behavioral Research, 31*(2), 239-270.

Monecke, A., & Leisch, F. (2012). semPLS: structural equation modeling using partial least squares.

Nitzl, C., Roldan, J. L., & Cepeda, G. (2016). Mediation analysis in partial least squares path modeling: Helping researchers discuss more sophisticated models. *Industrial management & data systems, 116*(9), 1849-1864.

Olsson, U. H., Foss, T., Troye, S. V., & 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, 7*(4), 557-595.

Reinartz, W., Haenlein, M., & 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.

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long range planning, 45*(5-6), 341-358.

Rigdon, E. E. (2016). Choosing PLS path modeling as analytical method in European management research: A realist perspective. *European Management Journal, 34*(6), 598-605.

Rigdon, E. E., Sarstedt, M., & Ringle, C. M. (2017). On comparing results from CB-SEM and PLS-SEM: five perspectives and five recommendations. *Marketing, 39*(3), 4-16.

Ringle, C. M., Sarstedt, M., & Straub, D. (2012). A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Quarterly (MISQ), 36*(1).

Sarstedt, M., & Mooi, E. (2019). Regression Analysis. In *A Concise Guide to Market Research* (pp. 209-256). Springer, Berlin, Heidelberg.

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research, 69*(10), 3998-4010.

Shmueli, G., Ray, S., Estrada, J. M. V., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research, 69*(10), 4552-4564.

Stone, M. (1974). Cross-validation and multinomial prediction. *Biometrika, 61*(3), 509- 515.

Sosik, J. J., Kahai, S. S., & 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 & Organization Management, 34*(1), 5-36.

Tenenhaus, M. (2008). Component-based structural equation modelling. *Total quality management, 19*(7-8), 871-886.

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational statistics & data analysis, 48*(1), 159-205.

Vinzi, V. E., Chin, W. W., Henseler, J., & Wang, H. (2010). *Handbook of partial least squares* (Vol. 201): Springer.

Wold, H. (1973). Nonlinear iterative partial least squares (NIPALS) modelling: some current developments *Multivariate Analysis–III* (pp. 383-407): Elsevier.

Wold, S. (1993). Discussion: PLS in chemical practice. *Technometrics, 35*(2), 136-139.