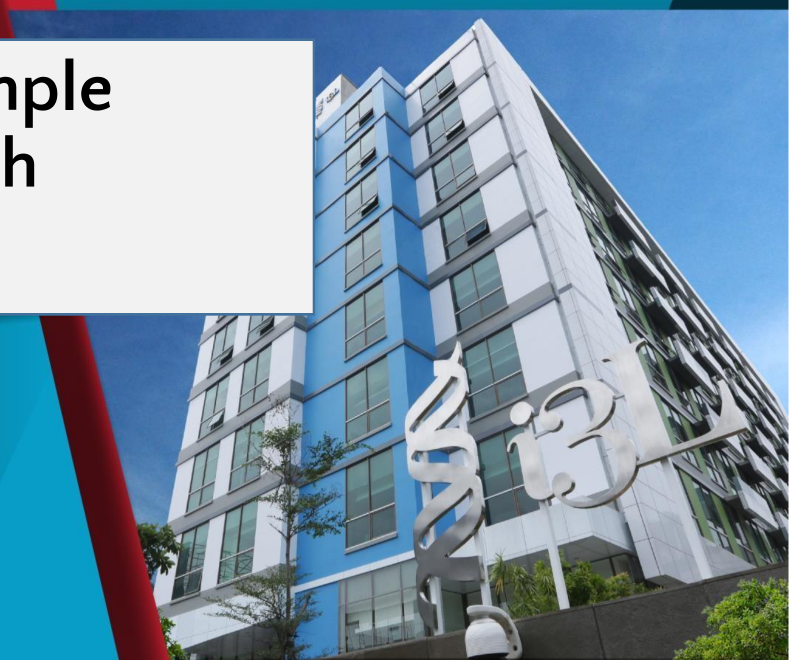
i3L Indonesia International Institute for Life Sciences

# Analyzing simple PLS-SEM with SmartPLS



i3L SCHOOL OF BUSINESS

# Introduction

This module will show you ho to use SmartPLS to analyze simple Partial Least Squares Structural Equation Modeling (PLS-SEM) quickly and easily. PLS-SEM is a multivariate analysis method that lets you explore complex relationships between observed and latent variables in a path model. It can handle different kinds of measurement models and effects, as well as various data issues. PLS-SEM has been widely applied in various disciplines such as marketing, management, psychology, education, and social sciences. This module is a concise guide that covers the essential topics and tips on how to use SmartPLS, one of the most popular software tools for PLS-SEM..

This module is organized into two parts.

- Part I explains the fundamentals of PLS-SEM.
- Part II demonstrates how to use SmartPLS to define measurement models, assess structural models, and interpret results.

By reading this module, you will learn how to conduct quantitative research analysis using PLS-SEM with SmartPLS in a fast and easy way.

Author: Lydia Karnadi.

Reference:


- Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019), “When to use and how to report the results of PLS-SEM”, European Business Review, Emerald Group Publishing Ltd., 14 January, doi: 10.1108/EBR-11-2018-0203.
- Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) Second Edition (Vol. 2).
- SmartPLS GmbH. (2023). Product | SmartPLS. Smartpls.com. https://www.smartpls.com/

# Table of Contents

# I. Partial Least Square - Structural Equation Modeling (PLS-SEM)

## 1.1. What is PLS-SEM and why use it?

Statistical analysis has become more important and complex for social science researchers due to the development of computer technology. Researchers need to use multivariate data analysis methods to understand more complicated relationships in their data.

PLS-SEM stands for **partial least squares structural equation modeling**, a component-based method for estimating complex cause-effect relationships in path models with latent variables. It is a **second-generation multivariate data analysis technique** in the family of **structural equation modeling** that differs from the covariance-based approaches (CB-SEM). Multivariate analysis is the application of statistical methodologies that analyze multiple variables simultaneously. PLS-SEM is primarily used to develop theories in **exploratory research** by focusing on explaining the variance in the dependent variables (Hair Jr. et al., 2017). For this reason, PLS-SEM is considered as a variance-based approach to SEM. It is important to note that PLS-SEM is not the same as PLS regression, another common multivariate.

PLS-SEM is suitable for **testing a theoretical framework from a prediction perspective** (Hair et. al, 2019), exploring theoretical extensions of established theories (exploratory research for theory development), or analyzing data with specific characteristics (e.g., financial ratios or archival data). PLS-SEM is not suitable for testing a theoretical framework from a confirmation perspective, testing hypotheses about mean differences or interaction effects, or analyzing data that fit a common factor model better than a composite model.

PLS-SEM has several advantages over other methods ((Hair et al., 2019), such as:

- It can handle **complex models** with many constructs, indicators and relationships.
- It **does not require normal data distribution** and can work with non-normal or skewed data (no distribution assumptions; PLS-SEM is a nonparametric method)
- It can accommodate **small sample sizes** as well as **large ones**, depending on the model complexity and the effect sizes.
- It can use **primary or secondary data**, which may lack a comprehensive substantiation on the grounds of measurement theory.
- It can estimate models with **formatively measured constructs**, which are measured by indicators that cause rather than reflect the construct.
- It can provide **latent variable scores** for follow-up analyses, such as cluster analysis or regression.

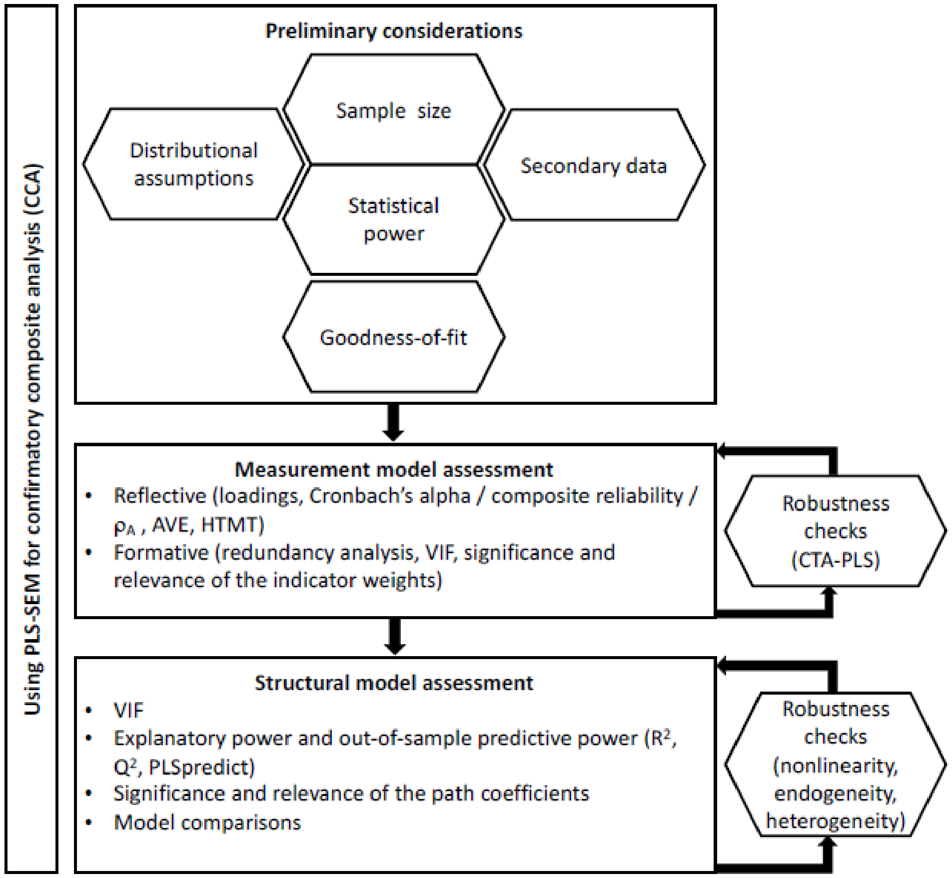



Figure 1.1.1.

Aspects and statistics to consider in a PLS-SEM analysis. Adapted from Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019), "When to use and how to report the results of PLS-SEM", European Business Review, Emerald Group Publishing Ltd., 14 January, doi: 10.1108/EBR-11-2018-0203.

## 1.2. Path models

Path models are diagrams used to visually display the hypotheses and variable relationships that are examined when SEM is applied, shown in figure 2 (Hair Jr. et al., 2017). Path models use circles or ovals for constructs (i.e., variables that are not directly measured) (Y1 to Y4) and rectangles for indicators (i.e., directly measured proxy variables) (x1 to x10). Arrows show the directional relationships between constructs and indicators. In PLS-SEM, single-headed arrows can be predictive or causal, if theory supports it.

A PLS path model has two elements: **a structural model** (or inner model in PLS-SEM) that shows the constructs and paths, and **measurement models** (or outer models in PLS-SEM) that show the constructs and indicators. In Figure 1.2.1., there are measurement models for **exogenous latent variables** (i.e., those explaining other constructs) and **endogenous latent variables** (i.e., those explained by other constructs). Researchers often refer to the measurement model of one latent variable instead of talking about measurement models of exogenous and endogenous latent variables, For example, x1 to x3 measure Y1 while x10 measures Y4.

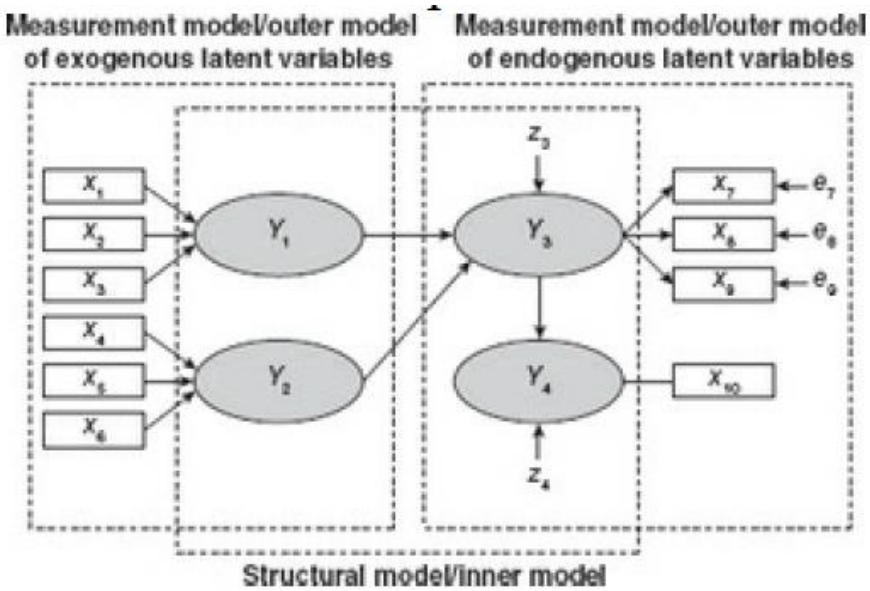



Figure 1.2.1.

A simple Path Model. Adapted from Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) Second Edition (Vol. 2).

## 1.3. Measurement Theory

Measurement theory defines how the latent variables (constructs) are measured. There are two ways to measure unobservable variables: **reflective and formative**. Constructs Y1 and Y2 in Figure 2 are measured formatively. The arrows go from the indicators (x1 to x3 for Y1 and x4 to x6 for Y2) to the construct, showing a causal (predictive) relationship. Y3 in the exhibit is measured reflectively. The arrows go from the construct to the indicators, showing that the construct causes the measurement (covariation) of the indicators. **Reflective measures have an error term** for each indicator, but **formative measures do not**. They are assumed to be error free. Y4 is measured with a single item, so the relationship between construct and indicator is **undirected**. The way to model constructs (i.e., formative vs. reflective and multi-items vs. single items) is important for developing path models.

## 1.4. Structural Theory

Structural theory shows how the latent variables are related (i.e., the constructs and paths in the structural model). The order and position of the constructs are based on theory or the researcher's knowledge. Path models go from left to right. The variables on the left are independent variables, and any variable on the right is the dependent variable. The variables on the left predict the variables on the right. But variables can also be both independent and dependent. When latent variables are only independent variables, they are exogenous latent variables (Y1 and Y2). When latent variables are only dependent variables (Y4) or both independent and dependent variables (Y3), they are endogenous latent variables. **Exogenous** latent variables have only single-headed arrows going out of them. **Endogenous** latent variables can have single-headed arrows going both into and out of them (Y3) or only going into them (Y4). The **exogenous latent variables Y1 and Y2 do not have error terms** because they explain the dependent variables in the path model.

## 1.5. Sample size in PLS-SEM & Statistical Power

PLS-SEM can handle **small sample sizes** with complex models that have many constructs and items (Hair et al., 2017). The PLS-SEM algorithm does this by computing measurement and structural model relationships separately with ordinary least squares regressions.

Although PLS-SEM can be used with smaller sample sizes, how small they are acceptable depends on the population size, the model structure, the significance level, and the effect sizes. Researchers should use **power analyses or power tables** (Hair et al., 2017) to determine the required sample size. Kock and Hadaya suggest two new methods for sample size calculations, which are the **inverse square root method** and **the gamma-exponential method**. Statistical analysis based on minimum sample size guidelines will ensure that the results of the statistical method are robust and that the model can be applied to another sample from the same population.

PLS-SEM can also analyze **large datasets**, including secondary data that may lack measurement theory support. CB-SEM and PLS-SEM yield very similar results with larger data sets (N = 250 or higher) when a sufficient number of indicator variables (four or more) are used to measure each construct (consistency at large).

When using PLS-SEM, account for **missing values** as in other statistical analyses. For up to 5% missing values per indicator, use mean replacement, EM (expectation-maximization algorithm), or nearest neighbor to get similar PLS-SEM estimates. Or delete observations with missing values, but this may reduce data variation and introduce bias.

### 1.5.1. The '10-times-rule' and Power Tables

A simple rule for estimating the sample size for PLS-SEM is to use the '10-times-rule' method (Hair, Jr et al., 2017). This means the sample size should be at least 10 times the maximum number of links pointing at any latent variable in the inner or outer model. Since sample size recommendations in PLS-SEM essentially build on the properties of OLS regression, researchers can follow Cohen's rules of thumb in his statistical power analyses for multiple regression, provided that the measurement models have acceptable quality in terms of outer loadings (factor loading should be above the common threshold of 0.70). Alternately, researchers should conduct individual power analyses, software such as G*Power (http://www.gpower.hhu.de/). A simple guideline for estimating the required sample size for PLS-SEM is to use power analyses based on the model part with the largest number of predictors.

Table 1.5.1.1

Sample size Recommendation in PLS-SEM for a Statistical Power of 80%. Adapted from Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) Second Edition (Vol. 2).

| Maximum Number of Arrows Pointing at a Construct | *Significance Level* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *1%* | | | | *5%* | | | | *10%* | | | |
| | *Minimum* $R^2$ | | | | *Minimum* $R^2$ | | | | *Minimum* $R^2$ | | | |
| | *0.10* | *0.25* | *0.50* | *0.75* | *0.10* | *0.25* | *0.50* | *0.75* | *0.10* | *0.25* | *0.50* | *0.75* |
| 2 | 158 | 75 | 47 | 38 | 110 | 52 | 33 | 26 | 88 | 41 | 26 | 21 |
| 3 | 176 | 84 | 53 | 42 | 124 | 59 | 38 | 30 | 100 | 48 | 30 | 25 |
| 4 | 191 | 91 | 58 | 46 | 137 | 65 | 42 | 33 | 111 | 53 | 34 | 27 |
| 5 | 205 | 98 | 62 | 50 | 147 | 70 | 45 | 36 | 120 | 58 | 37 | 30 |
| 6 | 217 | 103 | 66 | 53 | 157 | 75 | 48 | 39 | 128 | 62 | 40 | 32 |
| 7 | 228 | 109 | 69 | 56 | 166 | 80 | 51 | 41 | 136 | 66 | 42 | 35 |
| 8 | 238 | 114 | 73 | 59 | 174 | 84 | 54 | 44 | 143 | 69 | 45 | 37 |
| 9 | 247 | 119 | 76 | 62 | 181 | 88 | 57 | 46 | 150 | 73 | 47 | 39 |
| 10 | 256 | 123 | 79 | 64 | 189 | 91 | 59 | 48 | 156 | 76 | 49 | 41 |

Table 1.5.1.1. shows the minimum sample size for assuming the commonly used of 80% statistical power and 1%, 5%, and 10% confidence levels to detect R2 values of 0.10, 0.25, 0.50, and 0.75 in any endogenous construct in the structural model. The sample size depends on the model complexity (i.e., the maximum number of arrows pointing at a construct). For example, with five independent variables in the measurement and structural models, we need 45 observations to detect R2 values of at least 0.25 with 5% error probability. However, other researcher argue that the minimum sample size resulting from these calculations may still be too small.

### 1.5.2. Inverse square root method

Kock and Hadaya suggested the inverse square root method for minimum sample size in PLS path modeling (Hair, Jr et al., 2019). It is based on the probability of a path coefficient ratio exceeding a critical value at a given significance level. It only depends on one path coefficient and not on the model complexity. Assuming a power level of 80% and a significance level of 5%, the minimum sample size required for a minimum path coefficient of 0.2 is 155 observations.

## 1.6. Goodness of Fit

The purpose of **goodness of fit** is to measure how well a statistical model fits a set of observations. It summarizes the discrepancy between observed values and expected values under the model. It can be used to test hypotheses about the distribution of a population or the relationship between variables.

PLS-SEM is a method for structural equation modeling that estimates complex cause-effect relationships in path models with latent variables. Unlike CB-SEM, which relies on model fit, PLS-SEM focuses on prediction and theory testing. Since PLS-SEM lacks an established global measure of goodness-of-fit, its use for theory testing and confirmation is typically limited (Hair et al., 2017).

However, recent research has begun to develop goodness-of-fit measures within a PLS-SEM framework, thus expanding the method's applicability. In order to validate a model, Henseler et al. developed the **standardized root mean square residual (SRMR)**, which assesses the average magnitude of the discrepancies between observed and expected correlations with thresholds for assessing model fit < 0.08 (Hair et al., 2019). This measure has been incorporated into the SmartPLS 3 software as well.

Another type of measure is **NFI (normed fit index)**, which compares the discrepancy of the estimated model to that of a baseline model. NFI > 0.90 indicates a good fit, like SRMR. However, both types of measures are not well-established.

Therefore, researchers should be cautious when using these measures and validate their results. (Hair et al., 2019). For example:

1. They have not been thoroughly tested yet, so the suggested thresholds are uncertain.
2. They are not compatible with the PLS-SEM algorithm, which does not minimize the difference between observed and estimated covariances. Therefore, even bootstrap-based model fit assessments on the grounds of some distance measure or SRMR , which quantify the divergence between the observed and estimated covariance matrices may not work well.
3. 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., 2017).

# II. Analyze SmartPLS Result

## 2.1. What is SmartPLS?

SmartPLS is a software application for performing partial least squares structural equation modeling (PLS-SEM). PLS-SEM is a multivariate data analysis method that allows researchers to test complex relationships between latent variables and indicators. SmartPLS combines state-of-the-art methods (e.g., PLS-POS, IPMA, complex bootstrapping routines) with an easy-to-use and intuitive graphical user interface (SmartPLS GmbH, 2023). SmartPLS can be used for various purposes, such as prediction, theory testing, measurement model assessment, and data exploration

Measurement model validity and reliability are two criteria that can be used to evaluate the quality of the indicators and the constructs in a PLS-SEM model. **Validity** refers to the extent to which the indicators and the constructs measure what they are supposed to measure, while **reliability** refers to the extent to which the indicators and the constructs are consistent and accurate in their measurement. The summary can be seen below

Figure 2.1.1

Example of a Path Model With Three Constructs. . Adapted from Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). A Primer on Partial Least Squares Structural

| Criterion | Formative | Reflective |
|---|---|---|
| Causal priority between the indicator and the construct (Diamantopoulos and Winklhofer, 2001) | From the indicators to the construct | From the construct to the indicators |
| Is the construct a trait explaining the indicators or rather a combination of the indicators (Fornell and Bookstein, 1982) | If combination | If trait |
| Do the indictors represent consequences or causes of the construct?(Rossiter, 2002) | If causes | If consequences |
| Is it necessarily true that if the assessment of the trait changes, all items will change in a similar manner (asuuming they are equaly coded) (Chin, 1998) | No | Yes |
| Are the items mutually interchangeable (Jarvis, MacKenzie, and Podsakoff, 2003) | No | Yes |

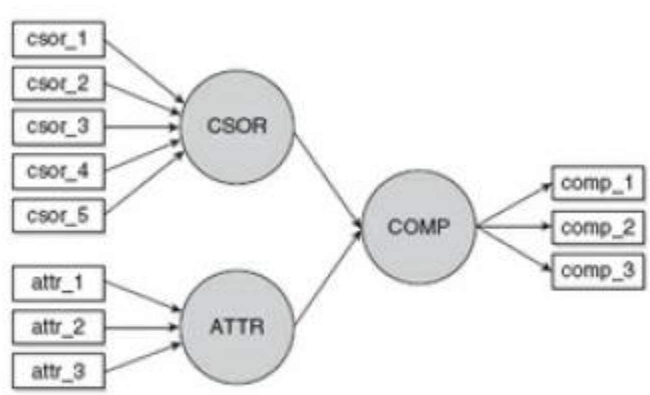


Figure 2.1.1

Example of a Path Model With Three Constructs. . Adapted from Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). A Primer on Partial Least Squares Structural

To assess measurement model validity and reliability, you need to distinguish between reflective and formative measurement models:

### Reflective measurement models

Reflective measurement models assume that the indicators are caused by the constructs. The reflective measurement model is based on classical test theory. It assumes that measures are the effects of an underlying construct. Therefore, causality goes from the construct to its measures. For example, when you examine the endogenous construct competence (COMP) in Figure 3.1.1., the direction of the arrows goes from the construct to the indicators (Comp_1, Comp_2, Comp_3). This type of measurement model is referred to as reflective.

Reflective indicators (or effect indicators) are a sample of all the possible items in the construct domain. Therefore, indicators of the same construct should be highly correlated and interchangeable, and removing an item should not change the construct's meaning, as long as the construct is reliable. If the latent trait evaluation changes (e.g., due to a different standard of comparison), all indicators will change together. A set of reflective measures is called a scale.

### Formative measurement models

Formative measurement models assume that the construct is formed by linear combinations of causal indicators. Therefore, causality goes from the indicators to the construct. For example, when you examine the two exogenous constructs – corporate social responsibility (CSOR) and attractiveness (ATTR) in Figure 2.1.1., the direction of the arrows is from the measured indicator variables to the constructs. This type of measurement model is called formative. Formative indicators (or causal indicators) are not interchangeable but capture specific aspects of the construct domain. The items define the construct, so removing an indicator may change the construct's nature. Therefore, covering the construct domain broadly is essential to capture the construct's content well.

## 2.2. How to define reflective measurement models?

To assess a reflective measurement model, you need to check four aspects (Hair et al., 2018):

1. indicator reliability
2. internal reliability
3. convergent validity, and
4. discriminant validity.

You can use SmartPLS to calculate these four criteria automatically or manually by running the **PLS algorithm.** The researcher may choose to delete some indicators from a construct to improve the four criteria above. However, this should be done with caution. Deleting indicators may enhance the reliability or discriminant validity, but it may also lower the content validity of the measurement.

## 2.2.1. Indicator reliability

Indicator reliability or indicator loadings show how well each indicator reflects the construct. They should be above 0.708, which means that the construct accounts for more than half of the indicator's variance.

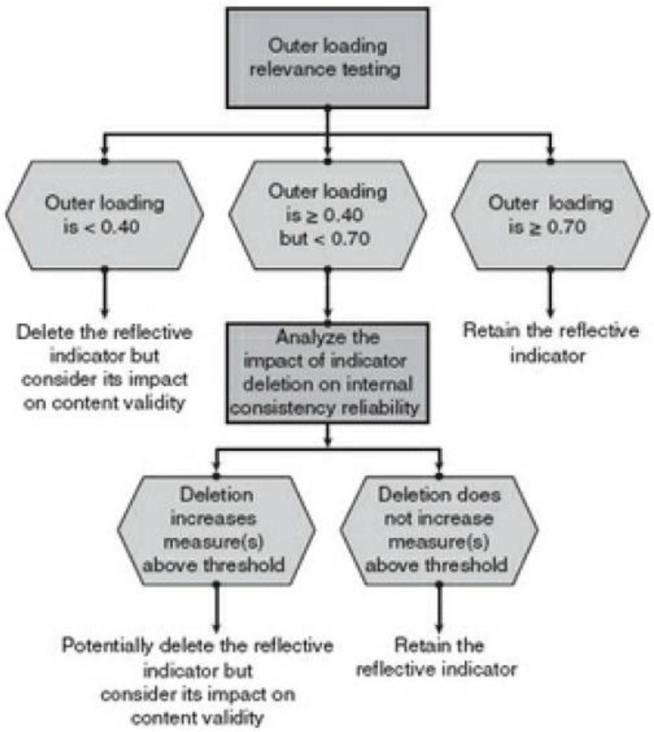



Figure 2.2.1.1.

Example of a Path Model With Three Constructs. . Adapted from Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). A Primer on Partial Least Squares Structural

To obtain weaker outer loadings (< 0.70) in social science studies is common, especially with new scales. Instead of deleting indicators below 0.70 automatically, researchers should check the effects of item removal on the composite reliability and the content validity of the construct.

- Indicators with outer loadings between 0.40 and 0.70 should be removed only when it increases the composite reliability (or the average variance extracted; see next section) above the threshold value.
- Content validity may also affect the decision to keep or delete an indicator with weaker outer loading. Indicators with very low outer loadings (below 0.40) should always be eliminated from the construct. According to Hair et al. (2017), content validity is assessed by evaluating whether the indicators of a construct adequately represent its domain. This can be done by using expert judgment, literature review, or empirical analysis

### 2.2.2. Internal reliability

**Internal reliability** or **internal consistency reliability** shows how well the indicators measure the same construct. The most common measures are:

- *Composite reliability*, which gives the upper bound or
- *Cronbach's alpha,* which gives the lower bound.

Higher values indicate higher reliability, but values above 0.95 may indicate redundancy or response bias. The ideal value is 0.80 to 0.90, but you can accept values as low as 0.70 or 0.60 in exploratory research.

### 2.2.3. Convergent validity

**Convergent validity** which is equivalent to the communality of the construct shows how much the construct explains the variance of its indicators. You can use average variance extracted (AVE) for this, which is the grand mean of the squared loadings of the indicators associated with the construct. The AVE should be at least 0.50, which means that the construct explains half or more of the variance of its indicators.

### 2.2.4. Discriminant validity

**Discriminant validity** shows how different the construct is from other constructs in the model. You can use different methods for this, such as Fornell-Larcker criterion, which compares the AVE of each construct with the squared correlation of that construct with other constructs. The AVE should be higher than the squared correlation for all constructs. However, this method may not be reliable in some cases, so you may need to use other methods such as heterotrait-monotrait ratio (HTMT) or cross-loadings.

#### *2.2.4.1. Fornell-Larcker criterion*

The Fornell-Larcker criterion compares the square root of the AVE of each construct with its highest correlation with any other construct in the model (same as comparing the AVE with the squared correlations). All reflective constructs must have higher square roots of the AVEs than their correlations with other latent variables, indicating valid and unique measures.

#### *2.2.4.2. Cross Loading*

Researchers have used two measures of discriminant validity in the past. The first one is the cross-loadings. This means that an indicator should load more on its own construct than on other constructs. The best way to show and report cross-loadings is in a table with indicators in rows and latent variables in columns. The loadings always surpass the cross-loadings.

#### *2.2.4.3. HTMT*

The HTMT criterion is more superior than the Fornell-Larcker criterion to assess discriminant validity. Henseler et al. (2015) show that the Fornell-Larcker criterion fails when indicators loading on a construct are similar (e.g., 0.65-0.85). They suggest the heterotrait-monotrait ratio (HTMT) of the correlations instead. The HTMT is the average item correlation between constructs over the average item correlation within constructs. High HTMT values imply discriminant validity problems.

Henseler et al. recommend a threshold value of 0.90 for very similar constructs, such as cognitive satisfaction, affective satisfaction, and loyalty. HTMT values above 0.90 indicate no discriminant validity. For more distinct constructs, a lower threshold value, such as 0.85, is advised (Hair et al., 2017).

## 2.3. How to define formative measurement models?

Formative constructs are best measured by PLS-SEM (Hair et al., 2018). To evaluate formative measurement models, we need to check:

(1) convergent validity,

(2) indicator collinearity, and

(3) statistical significance and relevance of the indicator weights

Researchers should be careful when deleting formative indicators based on statistics, for two reasons:

1. Formative indicator weights are affected by how many indicators are used to measure a construct. More indicators mean lower average weights. Therefore, formative models cannot have too many weights that are statistically significant.
2. Formative indicators should not be removed easily, because they are supposed to cover all aspects of the construct, as defined by the researcher in the conceptual stage. Unlike reflective models, formative indicators are not replaceable and removing even one indicator can reduce the content validity of the model.

**Convergent validity** means that the construct correlates highly with another measure of the same concept. This can be done by using redundancy analysis, which requires an alternative reflective measure of the construct in the research design. Cheah et al. suggest that a single-item measure is usually enough, but a multi-item measure may be better. If secondary data is used, a similar variable can be used as an alternative measure. The correlation between the formative and the alternative measure should be at least 0.70 (Hair et al., 2017).

**Indicator collinearity** can be assessed by the variance inflation factor (VIF), which should be low. High VIF values (above 5 or even 3) indicate problems with collinearity among the predictors.

**indicator weights for statistical significance and relevance** (i.e., size). Bootstrapping is used to test significance, since PLS-SEM is nonparametric. If the bootstrap distribution of the weights is skewed, BCa bootstrap confidence intervals should be used; otherwise, the percentile method is recommended (Hair et al., 2017). If an indicator weight has a confidence interval that includes zero, it may be removed from the model. However, the indicator's outer loading (i.e., its correlation with the construct) should also be considered. If both the weight and the loading are not significant, the indicator should be deleted. If the loading is significant but low (below 0.50), the indicator may be kept if there is strong theoretical support for it.

After checking the statistical significance of the indicator weights, researchers need to check each indicator's relevance. The indicator weights are standardized between -1 and +1, but sometimes they can be lower or higher than this, which is abnormal (e.g., due to collinearity and/or small sample sizes). A weight close to 0 means a weak relationship, while weights close to +1 (or -1) mean strong positive (or negative) relationships.

## 2.4. How to assess structural model?

After ensuring the reliability and validity of the construct measures, we examine the structural model results by developing **model fit** measures. Model fit indices help to judge how well a model structure matches the data and to find model errors. This includes checking the model's predictive capabilities and the construct relationships. However, we need to examine collinearity before assessing the structural relationships, to ensure that it does not bias the regression results. This process resembles assessing formative measurement models, but we use the latent variable scores of the exogenous constructs to calculate the VIF values (Hair et al., 2018). This analysis helps us to prevent common method bias (CMB) or the influence of common method variance (CMV) on the variables we measure.

Next, we consider the following main criteria to check the structural model (Hair et al., 2018):

1. The coefficient of determination ($R^2$)
2. The cross-validated redundancy measure $Q^2$, and
3. The statistical significance and relevance of the path coefficients.
4. The out-of-sample predictive power of their model by using PLSpredict if the sample size is large enough.

### 2.4.1. Collinearity Assessment (VIF)

The structural model coefficients are obtained from a series of regression equations. Before checking the structural relationships, collinearity must be checked to avoid biasing the regression results. This is similar to checking formative measurement models, but using the latent variable scores of the exogenous constructs to calculate the VIF values.

- High VIF values (above 5 or even 3) indicate collinearity problems among the predictors
- The VIF values should ideally be close to 3 or lower.

If collinearity is a problem, one option is to create higher order models that are supported by theory (Hair et al., 2017). If collinearity is not a problem, the next step is to look at the R2 value of the endogenous construct(s).

### 2.4.2. Coefficient of Determination ($R^2$ Value)

The $R^2$ or in-sample predictive power measures how much variance is explained in each of the endogenous constructs, and is therefore a measure of the model's explanatory power. The $R^2$ ranges from 0 to 1, with higher values indicating more explanatory power. As a guideline, $R^2$ values of 0.75, 0.50, and 0.25 can be seen as high, moderate, and low.

### 2.4.3. Blindfolding and predictive Relevance ($Q^2$ Value)

We can also assess the predictive accuracy of the PLS path model by calculating the $Q^2$ value. The $Q^2$ metric is based on the blindfolding procedure that removes, imputes with the mean, and estimates the model parameters for single points in the data matrix. The Q2 is not a measure of out-of-sample prediction, but rather a combination of out-of-sample prediction and in-sample explanatory power. Higher $Q^2$ values indicate higher predictive accuracy when the predicted and the original values are close. As a rule of thumb, $Q^2$ values above 0, 0.25, and 0.5 indicate small, medium, and large predictive relevance of the PLS-path model.

### 2.4.4. Structural Model Path Coefficients

After confirming the model's explanatory power and predictive power, the final step is to check the significance and relevance of the path coefficients. We interpret the path coefficients like the formative indicator weights. We run bootstrapping to check their significance and evaluate their values, which usually range from -1 to +1.

Estimated coefficients close to +1 mean strong positive relationships (and vice versa for negative values) that are usually significant (i.e., different from zero in the population). The closer the coefficients are to 0, the weaker the relationships. Very low values close to 0 are usually not significant. We use bootstrapping to get the standard error of each coefficient.

The bootstrapping routine is the same to check if a formative indicator contributes significantly to its construct. The bootstrap standard error lets us compute the empirical t values and p values for all path coefficients. When an empirical t value is bigger than the critical value, we say that the coefficient is significant at a certain error probability (i.e., significance level). In marketing, researchers usually use a significance level of 5%. Common critical values for two-tailed tests is 1.96 (significance level = 5%). Researchers usually report t values or p values to test the significance of all structural model relationships. We need to analyze the relative importance of relationships for interpreting the results and making conclusions

Most researchers use p values to check significance levels. A p value is the probability of wrongly rejecting a true null hypothesis (i.e., assuming a significant coefficient when it is not). When using a significance level of 5%, the p value must be less than 0.05 to say that the relationship is significant at a 5% level. For example, if we use a significance level of 5% and get a p value of 0.03 for a coefficient, we say that the coefficient is significant at a 5% level.

### 2.4.5. The out-of-sample predictive power of their model by using PLSpredict

The $R^2$ statistic is often seen as a measure of the model's predictive power. However, this is not completely correct, because the $R^2$ only shows the model's in-sample explanatory power—it does not say anything about the model's out-of-sample predictive power (Hair et al., 2018). To address this issue, Shmueli et al. 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 a holdout sample (Hair et al., 2018).

To assess the model's predictive power, we mainly focus on one key target construct (Hair et al., 2018). Where:

- We use RMSE for prediction error unless the distribution is very non-symmetric. Then, we use MAE.
- We check each indicator's Q2 predict value from PLS-SEM. Negative Q2 predict means no predictive power.
- We compare RMSE (or MAE) with LM for each indicator. We check if PLS-SEM has lower errors than LM for all (high predictive power), most (medium predictive power), some (low predictive power), or none (no predictive power) of the indicators.
- We check the error distribution. PLS-SEM residuals should be normal; left-tailed means over-prediction, right-tailed means under-prediction.
- We compare PLS-SEM and LM error distributions. They should be similar.