# 1 AN OVERVIEW OF RECENT AND EMERGING DEVELOPMENTS IN PLS-SEM



**LEARNING OUTCOMES**

1. Understand the origins and evolution of PLS-SEM.
2. Comprehend model specification in a PLS-SEM framework.
3. Describe the PLS-SEM algorithm's basic functioning principles.
4. Understand PLS-SEM's key characteristics vis-à-vis CB-SEM.

## CHAPTER PREVIEW

Along with the recent surge in applications of partial least squares structural equation modeling (PLS-SEM), methodological research has prompted numerous extensions of the original method that vastly increase its scope. In this chapter, we first provide an overview of the origins and evolution of PLS-SEM. This foundation will enable us to better understand why the method was slow to be adopted in the beginning, but has been increasingly applied in recent years across many social science disciplines, particularly in the various fields of business administration. We then discuss different aspects related to the specification of measurement and structural models, followed by a brief introduction of the PLS-SEM algorithm and selected extensions. Several considerations, which have their roots in the characteristics of the method, are important when applying PLS-SEM. We therefore discuss important characteristics of the PLS-SEM method that relate to the underlying measurement philosophy and the implications that arise from the way the algorithm estimates the model parameters. The chapter concludes with the introduction of a case study used throughout the remainder of this book.

## ORIGINS AND EVOLUTION OF PLS-SEM

The precursors to the PLS-SEM method were two iterative procedures that used least squares estimation to develop solutions for single and multicomponent models, and also for the method of canonical correlation (Wold, 1966). Further development of these procedures by Herman Wold led to the nonlinear iterative partial least squares (NIPALS) algorithm (Wold, 1973). A subsequent generalized version of the PLS-SEM algorithm focused on establishing and including latent variables in path models (Lohmöller, 1989, Chapter 2; Wold, 1980, 1982, 1985).

Several PLS methods evolved from Wold's generalized least squares algorithm (Mateos-Aparicio, 2011). One method is **principal components regression** (Hotelling, 1957; Jolliffe, 1982; Kendall, 1957; McCallum 1970), which performs a principal component analysis on the independent variables in which the model components are used as explanatory variables for a single dependent variable. However, principal components regression focuses on reducing the dimensionality of the independent variables only without considering the relationship between the independent and dependent variables.

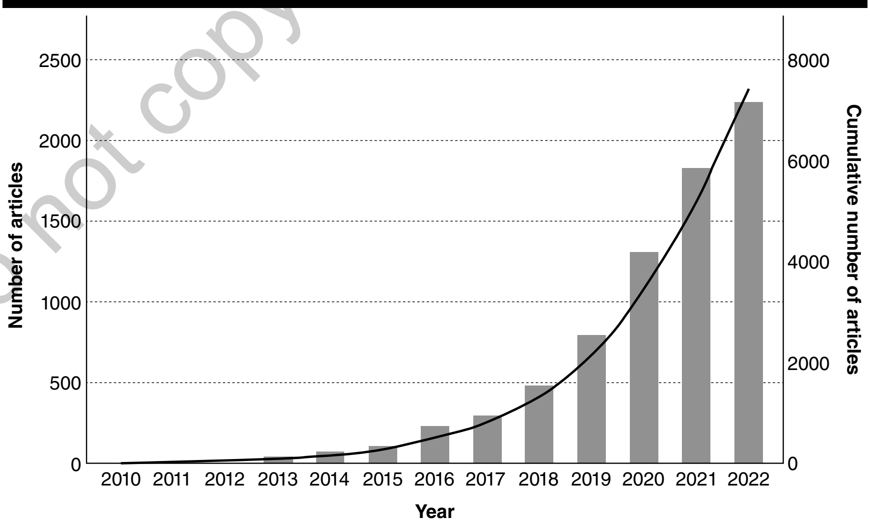
Another method is the **partial least squares regression (PLS-R)**, which was originally designed to reduce the issue of collinearity in regression models (Abdi, 2010; Kiers & Smilde, 2007; Wold, Ruhe, Wold, & Dunn, 1984). PLS-R focuses on the dimension reduction of the independent variables in a regression model intending to remove collinearity from the predictor variables. By doing so, the method optimizes the variance extracted from the independent variables while simultaneously maximizing the variance explained in the dependent variables. More precisely, PLS-R relies on a principal component analysis that extracts linear composites of the independent variables and their respective scores. Its aim is to reduce the dimensionality of the independent variables, while taking the relationship between the independent and dependent variables into consideration. As a result, PLS-R enables researchers to estimate models with more independent variables than observations in the dataset (Valencia & Diaz-Llanos, 2003).

Interestingly, PLS-R was not developed by Herman Wold but by his son, Svante Wold (e.g., Wold, Sjöström, & Eriksson, 2001), who worked in the field of analytical chemistry, known today as chemometrics. Together with Harald Martens, he adapted NIPALS to analyze chemical data. In addition to addressing the problem of multicollinearity in multiple regression models, their method solved the problem that arises when the number of variables is larger than the number of respondents (Martens, Martens, & Wold, 1983).

A third method that emerged from Wold's (1980) generalized PLS algorithm for estimating relationships between constructs and their indicators as well as between constructs (Lohmöller, 1989; Wold, 1982), also referred to as **partial least squares path modeling** (e.g., Esposito Vinzi, Trinchera, & Amato, 2010; Tenenhaus, Esposito Vinzi, Chatelin, & Lauro, 2005), the PLS approach to structural equation modeling (e.g., Chin, 1998), and **partial least squares structural equation modeling (PLS-SEM**; e.g., Hair, Ringle, & Sarstedt, 2011; Sarstedt, Hair, & Ringle, 2023). PLS-SEM determines the parameters of a set of equations in a path model by combining principal components analysis to assess the measurement models with path analysis to estimate the relationships between constructs (i.e., latent variables). Wold (1982, 1985) proposed his "soft modeling basic design" underlying PLS-SEM as an alternative to Jöreskog's (1973) **covariance-based structural equation modeling (CB-SEM)**. The alternative CB-SEM method has been labeled as hard modeling because of its more stringent assumptions in terms of data distribution and requiring much larger sample sizes (e.g., Falk & Miller, 1992). Importantly, for PLS-SEM "it is not the concepts nor the models nor the estimation techniques which are 'soft,' only the distributional assumptions" (Lohmöller, 1989, p. 64). While both methods were developed at about the same time, CB-SEM became much more widely applied because of its early availability through the LISREL software in the late 1970s. In contrast, the first software for PLS-SEM was LVPLS, which appeared in the mid-1980s (Lohmöller, 1984, 1987). But this initial software was not very user-friendly, and it was not until Chin's (1994) PLS-Graph software in the mid-1990s that PLS-SEM experienced wider application. With the release of SmartPLS 2 in 2005 (Ringle, Wende, & Will, 2005), SmartPLS 3 in 2015 (Ringle, Wende, & Becker, 2015), and SmartPLS 4 in 2022 (Ringle, Wende, & Becker, 2022), PLS-SEM applications increased exponentially (Sarstedt & Cheah, 2019), as evidenced by the popularity of the terms "PLS-SEM" and "PLS path modeling" in the Web of Science database in terms of articles published and citations (Exhibit 1.1).

Over the last two decades, there have been numerous introductory articles on the method (e.g., Chin, 1998; Haenlein & Kaplan, 2004; Hair, Risher, Sarstedt, & Ringle, 2019; Nitzl & Chin, 2017; Rigdon, 2013; Roldán & Sánchez-Franco, 2012;

**EXHIBIT 1.1 ■ Number of PLS-SEM-Related Articles per Year**




*Note:* Number of articles returned from the Web of Science database for the search terms "PLS-SEM" and "PLS path modeling" beginning from 2010.

Tenenhaus, Esposito Vinzi, Chatelin, & Lauro, 2005) as well as review articles examining how researchers in business and related fields have applied it (Exhibit 1.2). The usage of PLS-SEM also expanded into other research areas, such as agriculture, engineering, environmental sciences and ecology, geography, and psychology (Sarstedt, 2019).

With the increasing maturation of the PLS-SEM field (Hwang, Sarstedt, Cheah, & Ringle, 2020; Khan et al., 2019), researchers can draw on a much greater repertoire

**EXHIBIT 1.2 ■ Review Articles on PLS-SEM Usage**

| *Discipline* | *References* |
|---|---|
| Accounting | Lee, Petter, Fayard, & Robinson (2011)<br>Nitzl (2016) |
| Entrepreneurship | Manley, Hair, Williams, & McDowell (2020) |
| Family business | Sarstedt, Ringle, Smith, Reams, & Hair (2014) |
| Higher education | Ghasemy, Teeroovengadum, Becker, & Ringle (2020) |
| Hospitality and tourism | Ali, Rasoolimanesh, Sarstedt, Ringle, & Ryu (2018)<br>Do Valle & Assaker (2016)<br>Usakli & Kucukergin (2018) |
| Human resource management | Ringle, Sarstedt, Mitchell, & Gudergan (2020) |
| International business research | Richter, Sinkovics, Ringle, & Schlägel (2016) |
| Knowledge management | Cepeda-Carrion, Cegarra-Navarro, & Cillo (2019) |
| Management | Hair, Sarstedt, Pieper, & Ringle (2012) |
| Marketing | Hair, Sarstedt, Ringle, & Mena (2012)<br>Guenther, Ringle, Zaefarian, & Cartwright (2023)<br>Sarstedt, Hair, Pick, Liengaard, Radomir, & Ringle (2022) |
| Management information systems | Hair, Hollingsworth, Randolph, & Chong (2017)<br>Ringle, Sarstedt, & Straub (2012) |
| Operations management | Bayonne, Marin-Garcia, & Alfalla-Luque (2020)<br>Peng & Lai (2012) |
| Psychology | Willaby, Costa, Burns, MacCann, & Roberts (2015) |
| Quality management | Magno, Cassia, & Ringle (2022) |
| Software engineering | Russo & Stol (2021) |
| Supply chain management | Kaufmann & Gaeckler (2015) |

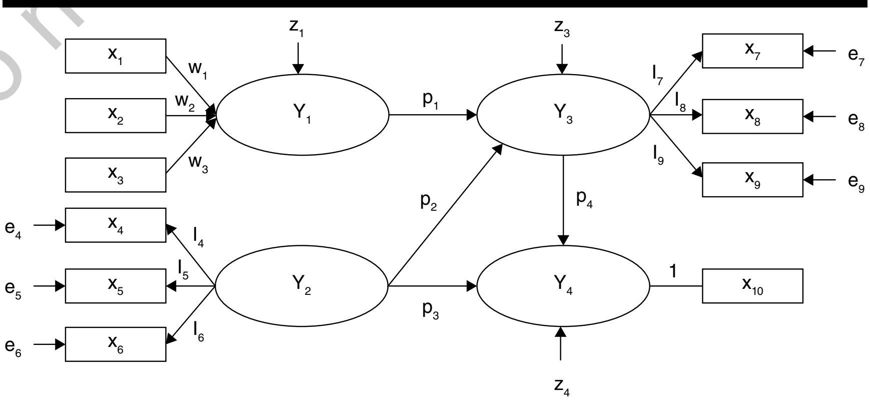
of advanced modeling, analysis techniques, and robustness checks (e.g., Sarstedt, Ringle, Cheah, Ting, Moisescu, & Radomir, 2020) to support their conclusions' validity and to identify more complex relationship patterns. For example, methodological research has made considerable progress in the treatment of observed heterogeneity in the context of moderator analysis (Becker et al., 2023; Memon et al., 2019), multigroup analysis (Chin & Dibbern, 2010; Klesel, Schuberth, Niehaves, & Henseler, 2021), and invariance assessment (Henseler, Ringle, & Sarstedt, 2016). Researchers have also developed novel latent class procedures (Becker, Rai, Ringle, & Völckner, 2013; Schlittgen, Ringle, Sarstedt, & Becker, 2016) and guidelines for their use (Sarstedt, Ringle, & Hair, 2021). Furthermore, progress has been made in the specification, estimation, and validation of higher order models (Becker, Cheah, Gholamzade, Ringle, & Sarstedt, 2023; Sarstedt, Hair, Cheah, Becker, & Ringle, 2019) and nonlinear effects (Basco, Hair, Ringle, & Sarstedt, 2021). With these extensions, PLS-SEM has become a full-fledged estimator for latent variable models and is capable of handling many modeling problems social sciences researchers face today.

## MODEL SPECIFICATION

Model specification in PLS-SEM involves two types of sub-models: the structural model and the measurement models. The **structural model** (also referred to as the **inner model**) specifies the relationships between the constructs. Constructs that act only as independent variables are referred to as **exogenous constructs**, whereas those that act as dependent variables are called **endogenous constructs**. The relationships between constructs are typically visualized in a path model, such as the one shown in Exhibit 1.3.

In this path model, $Y_3$ and $Y_4$ act as endogenous constructs, while $Y_1$ and $Y_2$ are exogenous. The endogenous constructs have **error terms** $z_3$ and $z_4$ connected to them,

**EXHIBIT 1.3 ■ Partial Least Squares Path Model Example**



which represent the unexplained variance (i.e., the difference between the model's in-sample prediction and actual construct scores). The exogenous construct $Y_1$ also has an error term ($z_1$) but in PLS-SEM, this error term is zero by default because of the way the method treats the measurement model of this particular construct, which is formative in nature (Diamantopoulos, 2011). Therefore, the error term is typically omitted in the display of a PLS path model. The other exogenous construct $Y_2$ is based on a reflective measurement model (i.e., the arrows point from the construct to its indicators) and therefore has no error term attached to it.

The structural model relationships in Exhibit 1.3 can be expressed using the following formulas:

$$Y_3 = p_1 \cdot Y_1 + p_2 \cdot Y_2 + z_3 \text{ and}$$
$$Y_4 = p_3 \cdot Y_2 + p_4 \cdot Y_3 + z_4$$

The **measurement models** (also referred to as **outer models**) express the relationships between each construct and its indicators. There are two broad ways to conceptualize measurement models from a measurement theory perspective. The first approach is referred to as reflective measurement. In a **reflective measurement model**, the indicators are considered to be error-prone manifestations of an underlying construct. That is, the relationships between the construct and the indicators are likely to include errors. Reflective measurement model relationships are represented as arrows going from the construct to its indicators. For example, the construct $Y_2$ in Exhibit 1.3 has a reflective measurement model, which can be expressed in the following way:

$$x_4 = l_4 \cdot Y_2 + e_4,$$
$$x_5 = l_5 \cdot Y_2 + e_5, \text{ and}$$
$$x_6 = l_6 \cdot Y_2 + e_6.$$

The terms $l_4$ to $l_6$ are the standardized indicator loadings, which are calculated from three bivariate regressions in which each reflective indicator $x_4$ to $x_6$ acts as a dependent variable and the construct $Y_2$ as independent variable (i.e., $l_4$ to $l_6$ simply represent the bivariate correlations between construct $Y_2$ and each of its indicators $x_4$ to $x_6$); $e_4$ to $e_6$ are the error terms representing the unexplained variance in each regression model. Note that there is no regression intercept, as PLS-SEM works with standardized data (i.e., the intercept is zero). **Reflective indicators** (sometimes referred to as **effect indicators** in the psychometric literature) can be viewed as a representative sample of all the possible items available in the conceptual domain of the construct (Nunnally & Bernstein, 1994). Since a reflective measurement model requires that all items reflect the same construct, indicators associated with a particular construct should be highly correlated with each other. In addition, individual items should be interchangeable, and any single item can generally be left out without changing the meaning of the construct, as long as the construct has sufficient reliability. The fact that the relationship goes from the construct to its indicators implies that if the evaluation of the latent trait changes (e.g., because of a change in the standard of comparison), all indicators will change simultaneously—at least to some extent.

The other type of measurement model is formative measurement. In a **formative measurement model**, the indicators form the construct using linear combinations. A change in an indicator's value due to, for example, a change in a respondent's assessment of the trait being captured by the indicator changes the value of the construct. That is, variation in the indicators precedes variation in the constuct (Borsboom, Mellenbergh, & van Heerden, 2003), which means that, by definition, constructs with a formative measurement model are inextricably tied to their measures (Diamantopoulos, 2006). Besides the difference in the relationship between indicator(s) and construct, formative measurement models do not require correlated indicators. Formative measurement model relationships are represented by arrows leading from the indicators to the construct.

Researchers distinguish between two types of indicators in the context of formative measurement: composite and causal indicators. **Composite indicators** largely correspond to the above definition of formative measurement models in that they are combined in a linear way to form a variate, which is also referred to as a composite in the context of SEM (Bollen, 2011; Bollen & Bauldry, 2011). More precisely, the indicators fully form the construct (i.e., the construct's $R^2$ value is 1.0), which means the construct has zero error. Composite indicators have often been used to measure **artifacts**, which can be understood as human-made concepts (Henseler, 2017). Examples of such artifacts in marketing include the retail price index or the marketing mix (Hair, Sarstedt, & Ringle, 2019). However, composite indicators can also be used to measure attitudes, perceptions, and behavioral intentions (Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016; Rossiter, 2011; Rossiter, 2016), provided the indicators have conceptual unity in accordance with a clear theoretical definition (Gilliam & Voss, 2013). The PLS-SEM algorithm relies solely on the concept of composite indicators because of the way the algorithm estimates formative measurement models (e.g., Diamantopoulos, 2011).

**Causal indicators** also form the construct, but this type of measurement acknowledges that it is unlikely that any set of causal indicators can fully capture every aspect of a latent phenomenon (Bollen & Diamantopoulos, 2017; Diamantopoulos & Winklhofer, 2001). Therefore, constructs measured with causal indicators have an error term, which is assumed to capture all the other causes of the latent variable not included in the model (Diamantopoulos, 2006). The use of causal indicators is prevalent in CB-SEM, which—at least in principle—allows for explicitly defining the error term of a formatively measured latent variable. However, the nature of this error term is ambiguous, as its magnitude partly depends on other constructs embedded in the model and their measurement quality (Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016).

The path model in Exhibit 1.3 has one formatively measured construct, $Y_1$, which PLS-SEM estimates using composite indicators. The corresponding measurement model can therefore be expressed as follows:

$$Y_1 = w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + z_1,$$

in which $z_1 = 0$.

Rather than using multiple items to measure a construct, researchers sometimes opt for a **single-item measurement**. PLS-SEM proves valuable in this respect, as the method does not encounter identification problems when using less than three items in a measurement model as it is the case with CB-SEM. Single items have practical advantages, such as ease of application, brevity, and lower costs associated with their use. Unlike long and complicated scales, which sometimes result in a lack of understanding and mental fatigue for respondents, single items promote higher response rates since the questions can be easily and quickly answered (Fuchs & Diamantopoulos, 2009; Sarstedt & Wilczynski, 2009). However, single-item measures do not offer more for less. For instance, when partitioning the data into groups, researchers have fewer options since scores from only a single variable are available to partition the data. Similarly, information is available from only a single measure instead of several measures when using imputation methods to deal with missing values.

More importantly, from a psychometric perspective, single-item measures do not allow for the removal of measurement error (as is the case with multiple items), which generally decreases the measure's reliability. Note that, contrary to commonly held beliefs, single-item reliability can be estimated (e.g., Cheah, Sarstedt, Ringle, Ramayah, & Ting, 2018; Loo, 2002; Wanous, Reichers, & Hudy, 1997). In addition, opting for single-item measures in most empirical settings is a risky decision when it comes to predictive validity considerations. Specifically, the set of circumstances that would favor the use of single-item over multi-item measures is very unlikely to be encountered in practice. Finally, social sciences scholars often include complex constructs in their theoretical models, such as trust, commitment, cooperation, and satisfaction. Using a single-item measure to represent such complex attitudinal or behavioral concepts can reduce the validity of the construct. According to guidelines by Diamantopoulos, Sarstedt, Fuchs, Kaiser, and Wilczynski (2012), single-item measures should be considered only in situations when (1) small sample sizes are present (i.e., $N < 50$), (2) path coefficients (i.e., the coefficients linking constructs in the structural model) of 0.30 and lower are expected, (3) items of the originating multi-item scale are highly homogeneous (i.e., inter-item correlations $> 0.80$, Cronbach's alpha $> 0.90$), and (4) the items are semantically redundant. Against this background, we generally advise against the use of single items for construct measurement. For further discussions on the efficacy of single-item measures, see, for example, Kamakura (2015) and Sarstedt, Diamantopoulos, Salzberger, and Baumgartner (2016).

## MODEL ESTIMATION

### The Original PLS-SEM Algorithm

With an adequate sample (see Hair, Hult, Ringle, & Sarstedt, 2022, Chapter 1, for further details), which meets the minum sample size, researchers can use the PLS-SEM method for the model estimation (see also Ringle, Sarstedt, Sinkovics, & Sinkovics, 2023). Model estimation in PLS-SEM draws on a three-stage approach that belongs to the family of (alternating) least squares algorithms (Mateos-Aparicio, 2011). Exhibit 1.4 illustrates the **PLS-SEM algorithm** as presented by Lohmöller (1989, Chapter 2). Henseler et al. (2012) offer a graphical illustration of its stages. The algorithm starts with an initialization stage in which it establishes preliminary construct scores. To compute these scores, the algorithm typically uses unit weights (i.e., 1) for all indicators in the measurement models (Hair et al., 2022).

Stage 1 of the PLS-SEM algorithm iteratively determines the inner weights (i.e., the path coefficients) and construct scores employing a four-step procedure. Step #1 uses the initial construct scores from the initialization of the algorithm to determine

**EXHIBIT 1.4 ■ The Basic PLS-SEM Algorithm (adapted from Lohmöller 1989, p. 29)**

| | |
|---|---|
| **Initialization** | |
| ***Stage 1:*** | **Iterative estimation of weights and construct scores**<br>**Starting at Step #4, repeat Steps #1 to #4 until convergence is obtained.** |
| #1 | Inner weights (here obtained by using the factor weighting scheme)<br>$b_{ji} = \begin{cases} cov(Y_j; Y_i) & \text{if } Y_j \text{ and } Y_i \text{ are adjacent} \\ 0 & 0 \text{ else} \end{cases}$ |
| #2 | Inside approximation<br>$\tilde{Y}_j := \sum_i b_{ji} Y_i$ |
| #3 | Outer weights; solve for<br>$\tilde{Y}_{jn} = \sum_{k_j} \tilde{w}_{k_j} x_{k_j n} + d_{jn}$ in a Mode A block<br>$x_{k_j n} = \tilde{w}_{k_j} \tilde{Y}_{jn} + e_{k_j n}$ in a Mode B block |
| #4 | Outside approximation<br>$Y_{jn} := \sum_{k_j} \tilde{w}_{k_j} x_{k_j n}$ |
| ***Stage 2:*** | **Estimation of outer weights, outer loadings, and path coefficients** |
| ***Stage 3:*** | **Estimation of location parameters** |

the inner weights $b_{ji}$ between the adjacent constructs $Y_j$ (i.e., the dependent one) and $Y_i$ (i.e., the independent one) in the structural model. Please note that we use an extended nomenclature compared to our discussion of model specification. For example, the inner weights $b_{ji}$, which are used as provisional estimates of the path coefficients, have two indices, $i$ and $j$, representing the independent construct and dependent construct of the corresponding relationship. The literature suggests different approaches to determining the inner weights (Chin, 1998; Lohmöller, 1989; Tenenhaus, Esposito Vinzi, Chatelin, & Lauro, 2005). In the **factor weighting scheme**, the inner weight corresponds to the covariance between $Y_j$ and $Y_i$ and is set to 0 if the constructs are unconnected. The **path weighting scheme** considers the direction of the inner model relationships (Lohmöller, 1989, Chapter 2). Chin (1998, p. 309) notes that the path weighting scheme "attempts to produce a component that can both ideally be predicted (as a predictand) and at the same time be a good predictor for subsequent dependent variables." As a result, the path weighting scheme leads to slightly higher $R^2$ values in the endogenous constructs compared to the other schemes and should therefore be preferred. In most instances, however, the choice of the inner weighting scheme has very little bearing on the results (Lohmöller, 1989, Chapter 2; Noonan & Wold, 1982).

Step #2, the inside approximation, computes proxies for all constructs $\widetilde{Y}_j$ by using the weighted sum of its adjacent constructs' scores $Y_i$. Then, for all the indicators in the measurement models, Step #3 computes new outer weights, which indicate the strength of the relationship between each construct $\widetilde{Y}_j$ and its corresponding indicators. To do so, the PLS-SEM algorithm uses two different estimation modes (Exhibit 1.4). When using **Mode A** (i.e., **correlation weights**), the bivariate correlation between each indicator and the construct determines the outer weights. In contrast, **Mode B** (i.e., **regression weights**) computes indicator weights by regressing each construct on its associated indicators.

By default, estimation of reflectively specified constructs draws on Mode A, whereas PLS-SEM uses Mode B for formatively specified constructs. However, Cho et al. (2023) show that this reflex-like use of Mode A and Mode B is not optimal when using PLS-SEM for prediction purposes. Their simulation study indicates that Mode A provides higher degrees of out-of-sample predictive power in situations commonly encountered in empirical research (see also Becker, Rai, & Rigdon, 2013). Exhibit 1.4 shows the formal representation of these two modes, where $x_{k_j n}$ represents the raw data for indicator $k$ ($k = 1, \ldots, K$) of construct $j$ ($j = 1, \ldots, J$) and observation $n$ ($n = 1, \ldots, N$); $\widetilde{Y}_{jn}$ are the construct scores from the inside approximation in Step #2, $\widetilde{w}_{k_j}$ are the outer weights from Step #3, $d_{jn}$ is the error term from a bivariate regression, and $e_{k_j n}$ is the error term from a multiple regression. The updated weights from Step #3 (i.e., $\widetilde{w}_{k_j}$) and the indicators (i.e., $x_{k_j n}$) are linearly combined to update the constructs scores (i.e., $Y_{jn}$) in Step #4 (outside approximation). Note that the PLS-SEM algorithm uses standardized data as input and always standardizes the generated construct scores in Step #2 and Step #4. After Step #4, a new iteration starts. The algorithm terminates when the weights obtained from Step #3 change marginally from one iteration to the next (typically $1{\cdot}10^{-7}$), or when the maximum number of iterations is achieved (typically 300).

Stages 2 and 3 use the final construct scores from Stage 1 as input for a series of ordinary least squares regressions. These regressions compute the final outer loadings, outer weights, and path coefficients as well as related elements such as indirect, and total effects, $R^2$ values of the endogenous constructs, and the indicator and construct correlations (Lohmöller, 1989, Chapter 2).

### The Weighted PLS-SEM Algorithm

An extension of the PLS-SEM approach is the **weighted PLS-SEM (WPLS)** algorithm (Becker & Ismail, 2016). This modified version of the original PLS-SEM algorithm enables researchers to match sample and population structure (Cheah, Roldán, Ciavolino, Ting, and Ramayah, 2021).

When estimating a PLS path model, researchers typically seek to draw inferences about the population of interest. An important requirement for such inferences is that the sample is representative of the population. Probability sampling methods, such as simple random sampling or cluster sampling, meet this requirement as every member of a population has an equal probability of being selected in the sample (Sarstedt & Mooi, 2019, Chapter 4). In the probability sampling case, every observation in the sample would have the same weight in the PLS-SEM analysis. In practice, however, population members are often not equally likely to be included in the sample, for example, because of the use of non-probability sampling methods, such as quota sampling, which is the norm for social sciences research. To adjust for resulting biases, researchers may use **sampling weights** (also referred to as post-stratification weights) that assign the observations different relevance in the parameter estimation process (Sarstedt, Bengart, Shaltoni, & Lehmann, 2018). For example, if a population consists of an equal share of males and females but the sample comprises 60% males and 40% females, sampling weights ensure that females are weighted more strongly than males in the parameter estimation.

Different from the original PLS-SEM algorithm, the WPLS algorithm considers sampling weights $v_n$ in the calculation of the mean, the variance, and covariance (correlation) of the construct scores in each iteration. For example, indicator standardization should rely on the weighted mean $\hat{\bar{x}}$ and weighted variance $\widehat{var(x)}$, defined as follows:

$$\hat{\bar{x}} = \frac{\sum_{n=1}^{N} v_n x_n}{\sum_{n=1}^{N} v_n}$$

$$\widehat{var(x)} = \frac{\sum_{n=1}^{N} v_n \left(x_n - \hat{\bar{x}}\right)^2}{\left(\sum_{n=0}^{N} v_n\right)},$$

whereby the hat symbolizes that these are weighted results. Similarly, the correlation weights used in Mode A draw on the weighted covariances, while the regression weights in Mode B and the inner model weights should use the weighted standardized data as input. For example, the inner model weights are given by:

$$b_{ji} = cor_v\left(Y_i\right)^{-1} cor_v\left(Y_i, \tilde{Y}_j\right).$$

The effect of these corrections is that all the calculations during the iterative PLS-SEM algorithm (e.g., path coefficient estimates) are weighted with the sampling weights while retaining all information from the original data set in the model. As a result, WPLS provides more accurate average population model parameter estimates than the basic PLS-SEM algorithm when appropriate sampling weights are available. Becker and Ismail (2016) and Cheah, Roldán, Ciavolino, Ting, and Ramayah (2021) provide more details on WPLS.

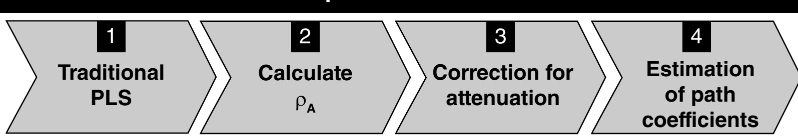
## Consistent PLS-SEM

The **consistent PLS-SEM (PLSc-SEM)** algorithm performs a correction of reflective constructs' correlations to make the results consistent with a common factor model (Dijkstra 2010, 2014; Dijkstra & Henseler 2015; Dijkstra & Schermelleh-Engel 2014). In principle, the correction builds on Nunnally's (1978) well-known correction for attenuation formula. As such, PLSc-SEM follows a composite modeling logic and modifies the PLS-SEM results to mimic a common factor model (Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016) and, thus, the results generated by CB-SEM (Jöreskog, 1978; Jöreskog & Wold, 1982).

Specifically, PLSc-SEM's objective is to compute the disattenuated (i.e., consistent) correlation $r^c_{Y_1,Y_2}$ between two constructs $Y_1$ and $Y_2$, which is the modified version of their original correlation ($r_{Y_1,Y_2}$), corrected for measurement error. To do so, PLSc-SEM divides the original construct correlation $r_{Y_1,Y_2}$ by the geometric mean of the constructs' reliabilities, measured using the **reliability coefficient $\rho_A$**.

PLSc-SEM follows a four-step approach (Exhibit 1.5). In Step 1, the traditional PLS-SEM algorithm is run. These results are then used in Step 2 to calculate the reliability $\rho_A$ of all reflectively measured constructs in the PLS path model (Dijkstra, 2014; Dijkstra & Henseler, 2015b). For formatively measured constructs and single-item constructs, $\rho_A$ is set to 1. In Step 3, the consistent reliabilities of all constructs from Step 2 are used to correct the inconsistent correlation matrix of the constructs

**EXHIBIT 1.5 ■ The Four-Step PLSc-SEM Procedure**



obtained in Step 1. More precisely, researchers obtain the consistent correlation between two constructs by dividing their correlation from Step 1 by the geometric mean (i.e., the square root of the product) of their reliabilities $\rho_A$. This correction applies to all correlations of reflectively measured constructs. The correlation of two formative and/or single-item constructs remains unchanged. The correction for attenuation only applies when at least one reflectively measured construct with a consistent reliability $\rho_A$ smaller than 1 is involved in the correlation between two constructs in the PLS path model. In Step 4, the consistent correlation matrix of the constructs allows re-estimating all model relationships yielding consistent path coefficients, corresponding $R^2$ values, and outer loadings. Note that significance testing in PLSc-SEM requires running an adjusted bootstrapping routine, which has also been implemented in extant PLS-SEM software.

In practical applications, PLSc-SEM results can be substantially influenced by low reliability levels of the constructs. As a result, the standardized path coefficients produced by PLSc-SEM can become very high (in some situations considerably larger than 1). Moreover, in more complex PLS path models, collinearity among the constructs has a strong negative impact on the PLSc-SEM results. In some instances, the structural model relationships become very small. Finally, bootstrapping of PLSc-SEM results frequently produces inadmissible solutions or can yield extreme outcomes, which result in high standard errors in certain relationships, increasing the Type II error rate.

In light of these limitations, the question arises as to when researchers should use PLSc-SEM. The PLSc-SEM approach is appropriate when researchers assume the data are obtained from a common factor model (Bollen, 2011; Bollen & Bauldry, 2011); we will discuss this model type in detail in the following section. In that case, the objective is to mimic CB-SEM results by assuming the construct can be adequately represented by the common variance of its indicators. Simulation studies for such models reveal that CB-SEM and PLSc-SEM return almost identical results of the estimated coefficients (Dijkstra & Henseler, 2015b). While CB-SEM and PLSc-SEM have approximately the same accuracy of estimated parameters and statistical power, PLSc-SEM retains most of PLS-SEM's advantageous features. Among others, PLSc-SEM does not rely on distributional assumptions, can handle complex models, is less affected by incorrect specifications in subparts of the model, and will not encounter convergence problems. At the same time, however, the correction for attenuation in Step 3 changes the path coefficient estimates, which have been derived from the original PLS-SEM estimation and maximize the endogenous constructs' explained variance. As a consequence, any assessment of the model's predictive power using the modified path coefficient estimates is inconsistent with the original PLS-SEM estimation. Considering that research has emphasized the causal-predictive nature as an integral part of PLS-SEM and a key distinguishing feature from CB-SEM, this limitation is highly problematic. In addition, Sarstedt, Hair, Ringle, Thiele, and Gudergan (2016) show that the bias produced by PLSc-SEM is considerably higher than the one produced by CB-SEM when erroneously using the method on data that stem from a composite model population.

In light of these limitations, Hair, Sarstedt, and Ringle (2019, p. 567) conclude that PLSc-SEM “adds very little to existing knowledge of SEM” and that researchers should revert to the widely recognized and accepted CB-SEM approach when estimating common factor models. Nevertheless, PLSc-SEM is an alternative to the standard CB-SEM estimation when attempting to estimate under-identified models or when convergence problems occur. The same limitations apply to Bentler & Huang’s (2014) PLSe2 method, which builds on the PLSc-SEM results but applies a generalized least squares covariance structure estimation on the modified correlation matrix, and Yuan, Wen, & Tang’s (2020) Cronbach’s α based approach. As such, the method does not unite the advantages of PLS-SEM and CB-SEM as suggested by some researchers (Ghasemy, Jamil, & Gaskin, 2021).

## PRINCIPLES OF PLS-SEM

Several considerations are important when applying PLS-SEM, which have their roots in the method’s characteristics. While PLS-SEM is a distinct statistical method (Schuberth, Zaza, & Henseler, 2023), some of its characteristics need to be explained by contrasting it to the CB-SEM method, which has long dominated social sciences research (Babin, Hair, & Boles, 2008). We first discuss the measurement philosophy underlying PLS-SEM, followed by aspects related to the way the method estimates model parameters and ensuing biases. We then summarize the implications regarding the recommended situations for the application of each method based on strengths and limitations.

### Philosophy of Measurement

A crucial conceptual difference between PLS-SEM and CB-SEM relates to the way each method calculates the scores of all constructs included in a model. CB-SEM represents a **common factor-based SEM** method that considers the constructs as common factors that explain the covariances between all indicators included in the theoretical model. The underlying assumption is that the covariances (or **common variance**) of a set of indicators can in principle, be perfectly explained by the existence of one unobserved variable (the common factor) and an individual random error (Spearman, 1927; Thurstone, 1947). The **common factor model** approach is consistent with the measurement philosophy underlying reflective measurement in which the indicators and their covariations are regarded as manifestations of the underlying construct. To estimate the model parameters, CB-SEM minimizes the divergence between the empirical (observed) covariance matrix and the covariance matrix implied (estimated) by the model given a certain set of parameter estimates (Hair, Black, Babin, & Anderson, 2019).

In contrast, PLS-SEM runs partial regressions to obtain construct scores that minimize the residuals (error variances) in the relationships between composites and indicators (i.e., in the measurement models) as well as those between composites (i.e., in the structural model)—see Tenenhaus, Esposito Vinzi, Chatelin, and Lauro (2005). In doing so, PLS-SEM linearly combines the indicators of each construct's measurement model to form composite variables. PLS is therefore considered a **composite-based SEM** method (Hwang, Sarstedt, Cheah, & Ringle, 2020). In estimating the composite scores, the PLS-SEM algorithm weights each indicator individually. The indicator weights reflect each indicator's importance in forming the composite. That is, indicators with a larger weight contribute more strongly to forming it. In addition, unlike CB-SEM, in which model solutions are based only on common variance, PLS-SEM solutions are derived from total variance, which consists of both common and unique variance.

Moreover, PLS-SEM composite scores are superior to **sum scores**, which unit-weight each item with a coefficient of 1 (or any other arbitrary value so long as it is constant) for representing constructs (McNeish & Wolf, 2020). For example, with reflective measurement models, the PLS-SEM weights are also indicative of each indicator's degree of measurement error. Indicators with high degrees of measurement error have smaller weights, thereby contributing less to forming the composite variable (Hair, Hult, Ringle, & Sarstedt, 2021). The process of applying weighted composites of indicator variables makes PLS-SEM superior to multiple regression and other statistical models using sum scores. If multiple regression with sum scores is used, the researcher assumes an equal weighting of all construct indicators, which means each indicator contributes equally to forming the composite, thereby not considering measurement error in calculating the construct scores (Hair & Sarstedt, 2019; Henseler et al., 2014). As a result, the use of sum scores produces higher parameter bias when the indicator weights differ, as is typically the case in applied research. Moreover, sum scores SEM approaches produce lower statistical power compared to differential indicator weights produced by PLS-SEM (Hair, Hult, Ringle, Sarstedt, & Thiele, 2017). McNeish and Wolf (2020) summarize the empirical and conceptual shortcomings of sum scores estimation, noting equal weights (1) generate unrealistic expectations about the population (data-generating) model by enforcing unnatural constraints on the empirical model; (2) hinder rigorous and accurate psychometric assessments by ignoring measurement theory in its entirety; (3) adversely affect construct validity and reliability; and (4) often result in vastly different conclusions due to inaccurate coefficient estimation. In addition, because virtually all the psychometric scales used in business research have been validated under the assumption of differentiated weights, using equal weights when applying these scales is categorically inappropriate because doing so imposes a different model from the initial validated one.

## Parameter Estimation Accuracy

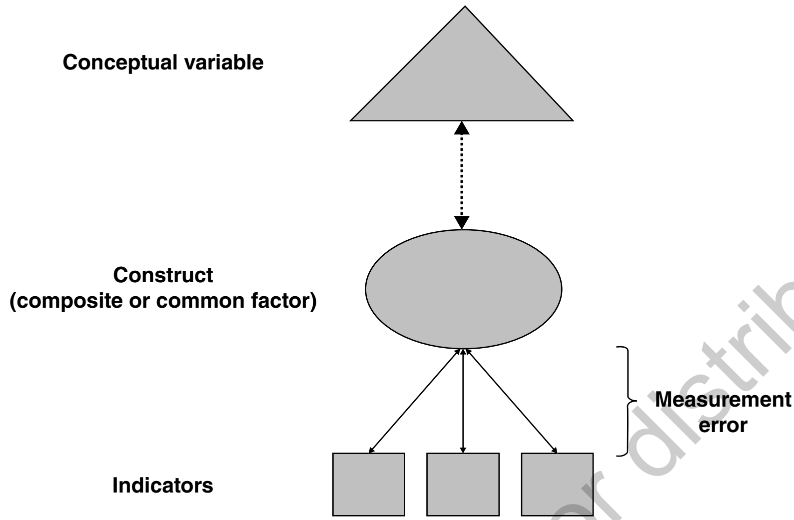
The philosophy of measurement assumed by PLS-SEM has important implications for any statement regarding its parameter accuracy (i.e., the degree to which the method produces accurate results). At the beginning of the method's development and use, researchers noted that PLS estimation is "deliberately approximate" (Hui & Wold 1982, p. 127) to common factor-based SEM, a characteristic that has come to be incorrectly known as the **PLS-SEM bias** (e.g., Marcoulides, Chin, & Saudners, 2012). Several studies have used simulations to demonstrate the alleged PLS-SEM bias (e.g., Goodhue, Lewis, & Thompson, 2012; McDonald, 1996; Rönkkö & Evermann, 2013), which supposedly manifests itself in measurement model estimates that are higher and structural model estimates that are lower compared to the prespecified values. The studies conclude that parameter estimates will approach what has been labeled the "true" parameter values when both the number of indicators per construct and sample size increase (Hui & Wold, 1982). All these simulation studies used CB-SEM as the benchmark against which the PLS-SEM estimates were evaluated based on the assumption that the results should be the same. However, such assessments can be expected to include bias since PLS-SEM is a composite-based approach that uses the total variance to estimate parameters (e.g., Schlittgen, Sarstedt, & Ringle, 2020; Schneeweiß, 1991; Tenenhaus, Esposito Vinzi, Chatelin, & Lauro, 2005). Not surprisingly, the very same issues apply when composite models are used to estimate CB-SEM results. In fact, Sarstedt, Hair, Ringle, Thiele, and Gudergan (2016) show that the biases produced by the CB-SEM methods are far more severe than those of PLS-SEM, when applying the method to the wrong type of model (i.e., estimating composite models with CB-SEM vs. estimating common factor models with PLS-SEM). Cho, Sarstedt, and Hwang (2022) recently confirmed these findings in a more complex simulation design and conceptually compared common factor and composite models, clarifying their similarities and differences. When acknowledging the different nature of the construct measures, most of the arguments voiced by critics of the PLS-SEM method (Rönkkö, McIntosh, Antonakis, & Edwards, 2016) are no longer an issue (Cook & Forzani, 2020; Cook & Forzani, 2023). Yuan and Fang (2022) raise additional conceptual concerns regarding the assumed PLS-SEM bias. They note that the parameter values in a CB-SEM analysis depend on the researcher's way of fixing the scale (e.g., fixing an indicator loading to unity)—a necessary step in any CB-SEM analysis. As a consequence, "the population values of the model parameters under [CB-]SEM are artificial," which implies that any resulting bias "does not enjoy a clear substantive interpretation." Cook and Foranzi (2023) recently echoed this notion.

Apart from these conceptual concerns, simulation studies show the differences between PLS-SEM and CB-SEM estimates when assuming the latter as a standard of comparison are very small, provided the measurement models meet minimum recommended standards in terms of measurement quality (i.e., reliability and validity). Specifically, when the measurement models have four or more indicators and the indicator loadings meet the common standards ($\geq 0.708$), there are practically no differences between the two methods in terms of parameter accuracy (e.g., Reinartz, Haenlein, & Henseler, 2009; Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016). Thus, the extensively discussed and supposed PLS-SEM bias is of no practical relevance for almost all SEM applications (e.g., Binz-Astrachan, Patel, & Wanzenried, 2014; Sharma, Liengaard, Sarstedt, Hair, & Ringle, 2023).

A more fundamental question is, however, whether it is reasonable to view common factor models as the universal measurement benchmark. Research casts considerable doubt on this premise, as common factors derived in CB-SEM are also not necessarily equivalent to the theoretical concepts that are the focus of research (Rigdon, 2012; Rigdon, Sarstedt, & Ringle, 2017; Rossiter, 2011; Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016). The reason for concern regarding the use of CB-SEM measurement and structural model results as a benchmark is that measurement in CB-SEM defines validity purely on the grounds of the relationships between the constructs and their indicators as measured by common variance only. For example, strong loadings in a measurement model suggest the construct converges into its indicators, offering evidence for convergent validity. However, constructs merely serve as representations or proxies of conceptual variables in a statistical model (Rigdon & Sarstedt, 2022; Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016). The actual entities of interest are the conceptual variables, which represent broad ideas or thoughts about abstract concepts that researchers establish and propose to measure in their research (e.g., Bollen, 2002). The relationships between the constructs and the conceptual variables, however, remain unknown since measurement error, as defined in SEM, only relates to the relationships between indicators and common factors or composites (Hair & Sarstedt, 2019). Exhibit 1.6 illustrates the relationships between conceptual variables, constructs, and indicators.

Even when considering perfect model fit, it is unreasonable to assume the constructs are equivalent to the conceptual variables they seek to represent (Cliff, 1983; MacCallum & Browne, 2007; Michell, 2013). The reason is any measurement includes **metrological uncertainty**, which refers to the dispersion of the measurement values attributed to the object or concept being measured (JCGM/WG1, 2008). Numerous sources contribute to metrological uncertainty, such as definitional uncertainty or limitations related to the measurement scale design, which go well beyond the simple standard errors produced by CB-SEM analyses (Hair & Sarstedt, 2019; Rigdon, Sarstedt, & Becker, 2020). One contributing factor to uncertainty is **factor (score) indeterminacy**, which means an infinite number of different sets of construct scores that will fit the model equally well are possible (Guttman, 1955; Schönemann & Wang, 1972). Factor indeterminacy produces a band of uncertainty in the relationship (i.e., the correlation) between a common factor inside the model and any variable outside the model (Steiger, 1979). Since the conceptual variable itself is not part of the model (i.e., it is outside the model), this range of correlation values applies as well to the relationship between an indeterminant common factor and the conceptual variable it is designed to represent (Exhibit 1.6). For example, the presence of factor indeterminacy implies the correlation between the common factor in the statistical model labeled, for example, customer satisfaction and the actual customer satisfaction is uncertain—even when the model shows a perfect fit (Rigdon & Sarstedt, 2022). To better grasp this concept, consider a model with four moderately correlated (0.3) common factors, each measured with three indicators and all with 0.7 loadings. In such a setting, the correlation between a common factor (e.g., the construct labeled *customer satisfaction*) and the corresponding

**EXHIBIT 1.6 ■ Error Framework for Conceptual Variables, Proxies, and Indicators**




conceptual variable (e.g., the actual *customer satisfaction*) lies in a range that has a width of 0.495 (Rigdon, Becker, & Sarstedt 2019). This means the range of uncertainty covers 0.495/2 = 24.75% of the total possible range of correlation (see Table 2 in Rigdon, Becker, & Sarstedt, 2019). Researchers' common practice of restricting the number of indicators per construct to improve model fit in CB-SEM analyses further increases this uncertainty (Hair, Matthews, Matthews, & Sarstedt, 2017; Rigdon, Becker, & Sarstedt, 2019).

While these issues do not necessarily imply that PLS-SEM is superior (Rigdon & Sarstedt, 2022), they cast considerable doubt on the assumption of some researchers that CB-SEM constitutes the gold standard when measuring unobservable concepts (e.g., Rhemtulla, van Bork, & Borsboom, 2020; Rigdon, 2016). Instead, researchers should recognize that all constructs are merely approximations of or proxies for conceptual variables (Exhibit 1.6), independent from how they were estimated (e.g., Rigdon, Sarstedt, & Ringle, 2017; Wickens, 1972). That is, constructs should be viewed as "something created from the empirical data which is intended to enable empirical testing of propositions regarding the concept" (Rigdon 2012, pp. 343–344). Hence, irrespective of the quality with which a conceptual variable is theoretically substantiated and operationally defined, and the rigor that encompasses measurement model development, any measurement in structural equation models produces only proxies for conceptual variables (Rigdon, 2012). This assessment is in line with the proliferation of all sorts of instruments that claim to measure essentially the same concept, although often with little chance to convert one instrument's measures into any other instrument's measures (Salzberger, Sarstedt, & Diamantopoulos, 2016). For example, research and practice have proposed a multitude of measurement instruments for corporate reputation, which rest on the same definition of the concept but differ fundamentally in terms of their underlying conceptualizations and measurement items (Sarstedt, Wilczynski, & Melewar, 2013). Similarly, Bergkvist and Langner (2017, 2019) find considerable heterogeneity in the operationalizations of common advertising constructs, such as attitude toward the ad, attitude toward the brand, ad credibility, ad irritation, and brand purchase intention. In addition, construct conceptualizations and operationalizations change over time (Bergkvist & Eisend, 2021), while the theoretical entity of interest (i.e., the conceptual variable) generally remains the same. These findings suggest there is no set way to perfectly measure a concept (Viswanathan, 2022). Nevertheless, much progress has been made in more accurately measuring concepts, particularly in consideration of multi-item versus single-item measures, and the addition of improved quantitative metrics for assessing both reliability and validity.

### Model Estimation Implications

An important characteristic of PLS-SEM is that the method does not simultaneously compute all the model relationships, but instead uses separate ordinary least squares regressions to estimate the model's partial regression relationships (Exhibit 1.4)—as implied by its name. As a result, the overall number of model parameters can be extremely high in relation to the sample size as long as each partial regression relationship draws on a sufficient number of observations (Chin & Newsted, 1999). For example, Antioco, Moenaert, Feinberg, and Wetzels (2008) estimate a PLS path model comprising 18 constructs and 33 structural model relationships with merely 121 observations. Needless to say, such an approach is unlikely to be reasonable from a sampling theory perspective unless the population of interest is highly homogeneous. Yet, it offers sufficient statistical power.

Upon convergence, the PLS-SEM algorithm produces a single specific (i.e., determinate) score for each observation per construct. Using these scores as input, PLS-SEM applies ordinary least squares regression with the objective of maximizing the explained variance values of the endogenous constructs and their indicators. That is, the method seeks to maximize **explanatory power**, also referred to as **in-sample predictive power**, which refers to the model's ability to reproduce the data that has been used to estimate the model parameters. Yuan and Fang (2022) have shown, numerically and by a simulation study, that PLS-SEM yields higher levels of explanatory power than CB-SEM. That is, when the aim is to maximize explanatory power, researchers should prefer PLS-SEM over CB-SEM.

The PLS-SEM results also enable assessment of the model's out-of-sample predictive power (or simply predictive power), which indicates a model's ability to predict new or future observations. As high explanatory power ($R^2$) does not guarantee significant predictive power (Inoue & Kilian, 2005; Sarstedt & Danks, 2022; Shmueli, 2010), researchers need to explicitly test this aspect of their model's performance using holdout samples

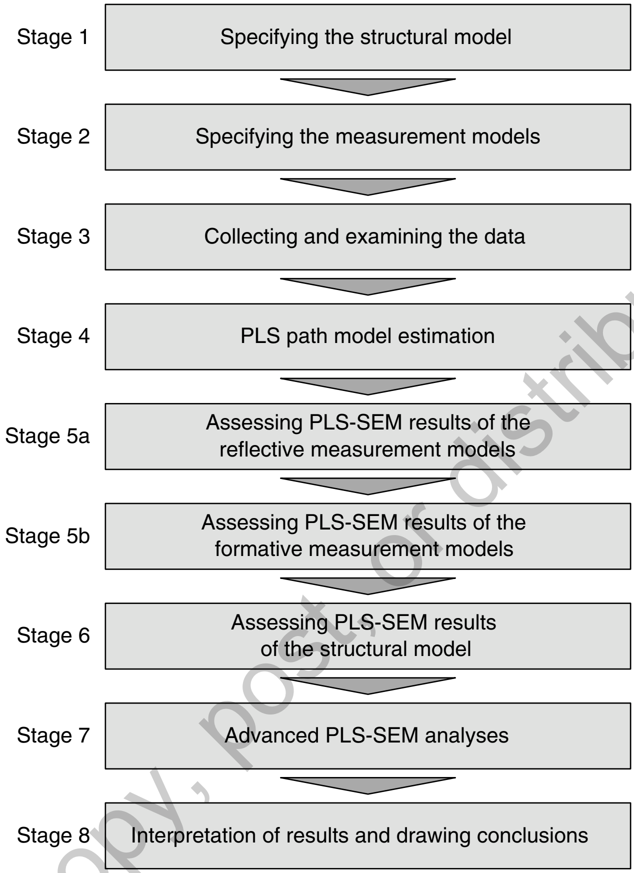
(Cepeda-Carrión, Henseler, Ringle, & Roldán, 2016), *k*-fold cross-validation (Shmueli, Ray, Velasquez Estrada, & Chatla, 2016; Shmueli et al., 2019) or a stand-alone predictive ability test (Sharma, Liengaard, Hair, Sarstedt, & Ringle, 2023). These characteristics define PLS-SEM's causal-predictive paradigm in which the aim is to assess the predictive power of a specified model carefully developed on the grounds of theory or logic. The underlying causal-predictive logic follows what Gregor (2006) refers to as **explaining and predicting (EP) theories**. EP theories imply an understanding of the underlying causes and prediction as well as a description of theoretical constructs and their relationships. According to Gregor (2006, p. 626), this type of theory "corresponds to commonly held views of theory in both the natural and social sciences." Numerous seminal theories and models, such as Oliver's (1980) expectation-disconfirmation theory or the various technology acceptance models (e.g., Venkatesh, Morris, Davis, & Davis, 2003) follow an EP-theoretic approach in that their aim is to explain *and* predict. PLS-SEM is perfectly suited to investigate models derived from EP theories as the method strikes a balance between machine learning methods, which are fully predictive in nature, and CB-SEM, which focuses on confirmation and model fit (Richter, Cepeda Carrión, Roldán, & Ringle, 2016). Its causal-predictive nature makes PLS-SEM particularly appealing for research in fields that aim to derive recommendations for practice (Chin et al., 2020; Sarstedt & Danks, 2022). For example, recommendations in managerial implications sections that are an element of many business research journals always include predictive statements ("our results suggest that managers should…"). Making such statements requires a prediction focus in model estimation and evaluation. PLS-SEM perfectly emphasizes this need as the method sheds light on the mechanisms (i.e., the structural model relationships) through which the predictions are generated (Hair, 2021; Hair & Sarstedt, 2021; Legate, Hair, Lambert, & Risher, 2021).

## ORGANIZATION OF THE REMAINING CHAPTERS

The remaining chapters provide more detailed information on advanced analyses using PLS-SEM, including specific examples of how to use the SmartPLS 4 software. In this book on advanced PLS-SEM issues, we build upon Stage 7 of the systematic procedure for applying PLS-SEM (Exhibit 1.7) from the *Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (Hair, Hult, Ringle, & Sarstedt, 2022), which enjoys an extremely high level of popularity (Ringle, 2019). The advanced PLS-SEM analyses will enable you to better understand and explain your results, and provide the types of analyses and diagnostic metrics editors and reviewers increasingly request. Exhibit 1.8 displays the chapters and topics covered in this book.

Chapter 2 introduces higher-order constructs, which allow measuring a conceptual variable on different levels of abstraction. For this purpose, a higher-order construct simultaneously models several subcomponents that cover more concrete traits of the conceptual variable represented by this construct. With the growing complexity

**EXHIBIT 1.7 ■ A Systematic Procedure for Applying PLS-SEM**




of theories and cause-effect models in the social sciences, researchers have increasingly used these models in their PLS-SEM studies (e.g., Sarstedt, et al., 2022).

Chapter 3 starts with an introduction to nonlinear relationships, which have gained increasing prominence in applications of PLS-SEM (Basco, Hair, Ringle, & Sarstedt, 2021). When the relationship between two constructs is nonlinear, the size of the effect between them depends not only on the magnitude of change in the exogenous construct but also on its value. We introduce the principles of nonlinear modeling and describe how to run corresponding analyses in SmartPLS 4. The second part of the chapter introduces confirmatory tetrad analysis (CTA-PLS), which facilitates empirically assessing whether the data support a formative or a reflective measurement model specification. In light of the potential biases that result from misspecifying measurement models, the CTA-PLS offers a valuable means to safeguard the results' validity.

Chapter 4 introduces two techniques that extend standard PLS-SEM results assessment procedures. We first introduce the necessary condition analysis (NCA) that

**EXHIBIT 1.8 ■ Thematical Vverview of this Book**

| Chapter | Topics |
|---|---|
| 2 | Higher-Order Constructs |
| 3 | Advanced Modeling and Model Assessment<br>— Nonlinear Relationships<br>— Confirmatory Tetrad Analysis (CTA-PLS) |
| 4 | Advanced Results Illustration<br>— Necessary Condition Analysis (NCA)<br>— Importance-Performance Map Analysis (IPMA) |
| 5 | Modeling Observed Heterogeneity<br>— Measurement Invariance Assessment (MICOM)<br>— Multigroup Analysis |
| 6 | Modeling Unobserved Heterogeneity<br>— Finite Mixture PLS (FIMIX-PLS)<br>— PLS Prediction-Oriented Segmentation (PLS-POS) |



takes a different perspective on the model relationships by testing the degree to which an outcome—or a certain level of an outcome—depends on the values of other constructs or indicators in the model. By assuming such necessity relationships, the NCA complements the sufficiency logic that standard PLS-SEM analyses rely on. In the second part of this chapter, we present the importance performance map analysis (IPMA). The IPMA contrasts the structural model relationships, which represent a construct's or an indicator's "importance" for a target construct with the average construct scores, which represent their "performance" in the PLS path model.

The following two chapters deal with different concepts that enable researchers to model heterogeneous data. Chapter 5 provides an overview of observed and unobserved heterogeneity, showing how disregarding heterogeneous data structures can generate biased results. Next, we discuss measurement invariance, which is a primary concern before comparing groups of sample data. The chapter concludes with an introduction of different types of multigroup analysis that are used to compare parameters (usually path coefficients) between two or more groups of data. While these methods enable researchers to account for observed heterogeneity, more often than not, situations arise in which differences related to unobserved heterogeneity prevent the derivation of accurate results as the analysis on the aggregate data level masks group-specific effects. Chapter 6 introduces two additional methods, finite mixture PLS (FIMIX-PLS) and prediction-oriented segmentation in PLS (PLS-POS) that enable researchers to identify and treat unobserved heterogeneity in PLS path models.

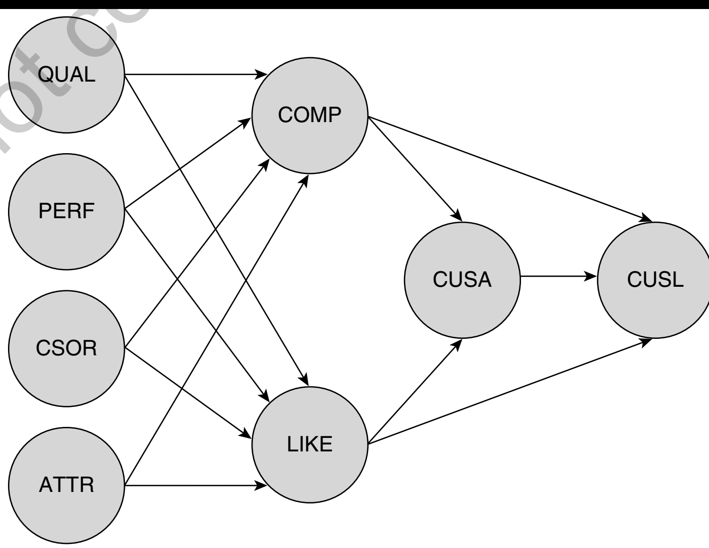
## CASE STUDY ILLUSTRATION

### Corporate Reputation Model

The most effective way to learn how to use a statistical method is to apply it to a set of data. Throughout this book, we use a single example that enables us to do that. The example is drawn from a series of published studies on corporate reputation, which is general enough to be understood by researchers from various disciplines, thus further facilitating comprehension of the analyses presented in this book. More precisely, we draw on the corporate reputation model by Eberl (2010), which Hair, Hult, Ringle, and Sarstedt (2022) use in their *Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. The model's purpose is to explain the effects of corporate reputation—a company's overall evaluation by its stakeholders (Helm, Eggert, & Garnefeld, 2010)—on customer satisfaction (*CUSA*) and, ultimately, customer loyalty (*CUSL*). Following Schwaiger (2004), corporate reputation is measured using two dimensions. One dimension represents the cognitive evaluations of the company and measures the construct describing the company's competence (*COMP*). The second dimension captures affective judgments and assesses perceptions of the company's likeability (*LIKE*). These two constructs are hypothesized to explain variations in customer satisfaction and loyalty (Schwaiger, Witmaier, Morath, & Hufnagel, 2021). Schwaiger (2004) further identifies four antecedent dimensions of reputation: quality (*QUAL*), performance (*PERF*), attractiveness (*ATTR*), and corporate social responsibility (*CSOR*). Exhibit 1.9 shows the corporate reputation model.

The measurement models of the *LIKE, COMP*, and *CUSL* constructs have three reflective indicators, whereas *CUSA* is measured with a single item. In general, we

**EXHIBIT 1.9 ■ Structural Model of Corporate Reputation**



recommend that using single items should be avoided, particularly in PLS-SEM analyses (e.g., Diamantopoulos, Sarstedt, Fuchs, Kaiser, & Wilczynski, 2012; Sarstedt, Diamantopoulos, & Salzberger, 2016; Sarstedt, Diamantopoulos, Salzberger, & Baumgartner, 2016). This single-item measure was included in our example, however, for illustrative purposes of how such a question could be included in obtaining a PLS-SEM solution. Finally, the four exogenous constructs are measured by a total of 21 formative indicators. Exhibit 1.10 provides an overview of all items and item wordings. Respondents rated the questions on 7-point Likert scales, with higher scores denoting higher levels of agreement with a particular statement. In the case of the *CUSA* indicator, higher scores denote higher levels of satisfaction (1 = *very dissatisfied;* 7 = *very satisfied*). Satisfaction and loyalty were measured with respect to respondents' own mobile phone service providers.

The measurement approach has been validated in different countries and applied in various research studies (e.g., Eberl & Schwaiger, 2005; Hult, Hair, Proksch, Ringle, Sarstedt, & Pinkwart, 2018; Radomir & Wilson, 2018; Raithel & Schwaiger, 2014; Raithel, Wilczynski, Schloderer, & Schwaiger, 2010; Schwaiger, Witmaier, Morath,

**EXHIBIT 1.10 ■ Overview of Constructs and Indicators**

| | |
|---|---|
| **Competence (*COMP*)** | |
| *comp_1* | [The company] is a top competitor in its market. |
| *comp_2* | As far as I know, [the company] is recognized worldwide. |
| *comp_3* | I believe that [the company] performs at a premium level. |
| **Likeability (*LIKE*)** | |
| *like_1* | [The company] is a company that I can better identify with than other companies. |
| *like_2* | [The company] is a company that I would regret more not having if it no longer existed than I would other companies. |
| *like_3* | I regard [the company] as a likable company. |
| **Customer Loyalty (*CUSL*)** | |
| *cusl_1* | I would recommend [the company] to friends and relatives. |
| *cusl_2* | If I had to choose again, I would choose [the company]. |
| *cusl_3* | I will remain a customer of [the company] in the future. |
| **Customer Satisfaction (*CUSA*)** | |
| *cusa* | If you consider your experiences with [the company], how satisfied are you with [the company]? |
| **Quality (*QUAL*)** | |
| *qual_1* | The products/services offered by [the company] are of high quality. |

**EXHIBIT 1.10 ■ Overview of Constructs and Indicators (Continued)**

| | |
|---|---|
| *qual_2* | [The company] is an innovator, rather than an imitator with respect to [the industry]. |
| *qual_3* | [The company]'s products/services offer good value for money. |
| *qual_4* | The services [the company] offered are good. |
| *qual_5* | Customer concerns are held in high regard at [the company]. |
| *qual_6* | [The company] is a reliable partner for customers. |
| *qual_7* | [The company] is a trustworthy company. |
| *qual_8* | I have a lot of respect for [the company]. |
| **Performance (*PERF*)** | |
| *perf_1* | [The company] is a very well-managed company. |
| *perf_2* | [The company] is an economically stable company. |
| *perf_3* | The business risk for [the company] is modest compared to its competitors. |
| *perf_4* | [The company] has growth potential. |
| *perf_5* | [The company] has a clear vision about the future of the company. |
| **Corporate Social Responsibility (*CSOR*)** | |
| *csor_1* | [The company] behaves in a socially conscious way. |
| *csor_2* | [The company] is forthright in giving information to the public. |
| *csor_3* | [The company] has a fair attitude toward competitors. |
| *csor_4* | [The company] is concerned about the preservation of the environment. |
| *csor_5* | [The company] is not only concerned about profits. |
| **Attractiveness (*ATTR*)** | |
| *attr_1* | [The company] is successful in attracting high-quality employees. |
| *attr_2* | I could see myself working at [the company]. |
| *attr_3* | I like the physical appearance of [the company] (company logo, buildings, shops, etc.). |

& Hufnagel, 2021; Schloderer, Sarstedt, & Ringle, 2014; Sharma, Shmueli, Sarstedt, Danks, & Ray, 2021). Research has shown that, compared to alternative reputation measures, Schwaiger's (2004) approach performs favorably in terms of convergent validity and predictive validity (Sarstedt, Wilczynski, & Melewar, 2013). The data set used for all analyses in this book stems from Hair, Hult, Ringle, and Sarstedt (2022) and has 344 responses regarding four major mobile network providers in Germany's mobile communications market (for a newer dataset, see also Sarstedt, Ringle, & Iuklanov, 2023).

## PLS-SEM Software

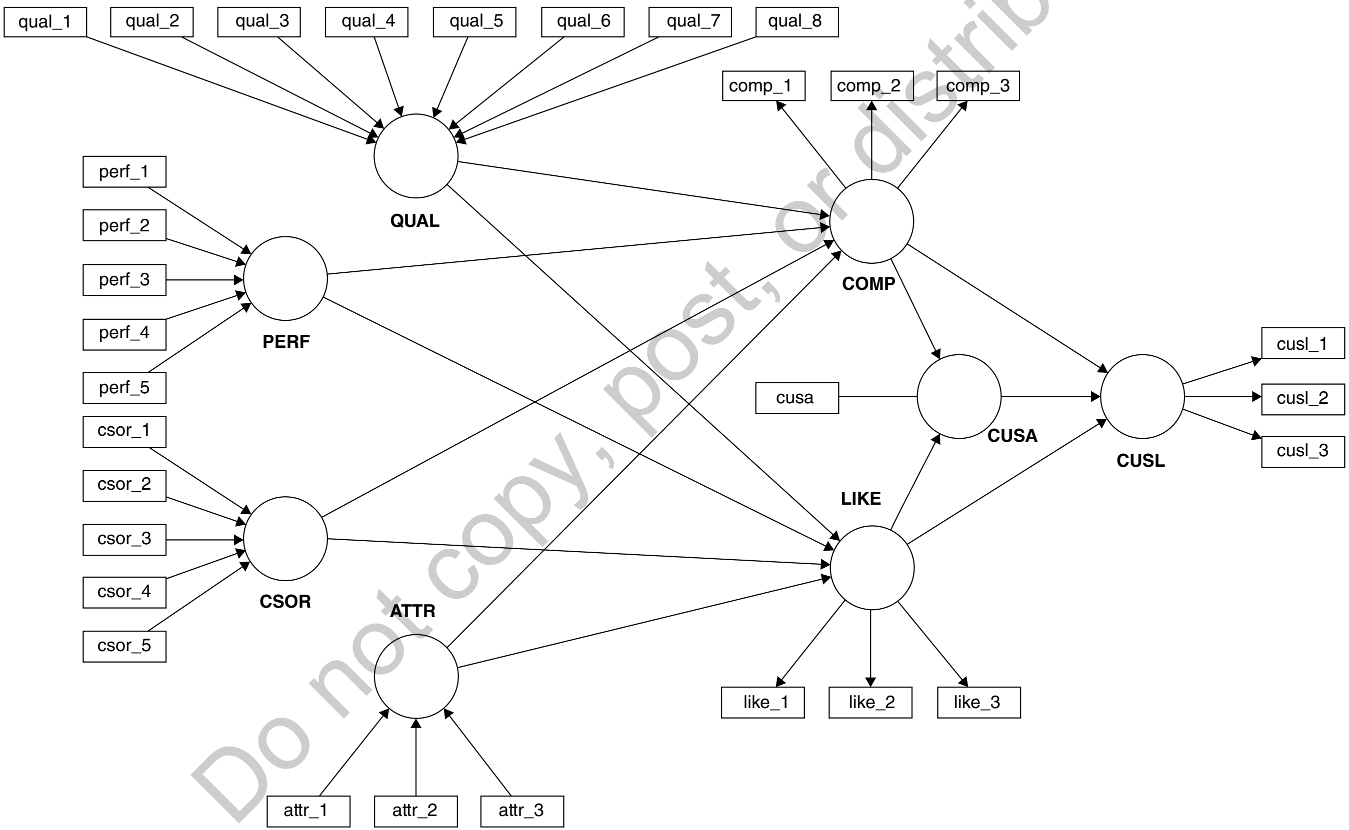
To establish and estimate PLS path models, users can choose from a range of software programs. A popular early example of a PLS-SEM software program is PLS-Graph (Chin, 1994), which is a graphical interface to Lohmöller's (1987) LVPLS, the first user-friendly PLS software. Compared to the original LVPLS, which required the user to enter commands via a text editor, PLS-Graph represents a significant improvement, especially in terms of user-friendliness. With the increasing dissemination of PLS-SEM in a variety of disciplines, several other programs with user-friendly graphical interfaces were introduced to the market, such as SmartPLS (Ringle, Wende, & Becker, 2022) and WarpPLS (Kock, 2020). Finally, users with experience in the statistical software environment R can also draw on packages, such as csem (Rademaker et al., 2021) and SEMinR (Ray et al., 2022), which facilitate the flexible analysis of PLS path models. The new R software workbook of the primer on PLS-SEM (Hair et al., 2022), with an electronic copy download available for free, illustrates all elements of the corporate reputation case study using the SEMinR package.

To date, SmartPLS is the most comprehensive and advanced program in the field (Sarstedt & Cheah, 2019). The software's most recent version, SmartPLS 4, therefore serves as the basis for all case study examples in this book. The student version of the software is available free of charge at https://www.smartpls.com. It offers practically all functionalities of the full version but is restricted to data sets with a maximum of 100 observations. However, as the data set used in this book has more than 100 observations (344 to be precise), we use the professional version of SmartPLS 4, which is available as a 30-day trial version at https://www.smartpls.com. The SmartPLS website includes many additional resources, such as short explanations of PLS-SEM and software-related topics, a list of recommended literature, answers to frequently asked questions, tutorial videos for getting started using the software, and the SmartPLS forum, which enables you to discuss PLS-SEM topics and share ideas with other users.

## Setting Up the Model in SmartPLS

Before we specify our model in SmartPLS 4, we need to have data that serve as the basis for running the model. SmartPLS 4 supports data imported from various file formats, such as Microsoft Excel (.xlxs), SPSS (.sav), comma-separated values (.csv), and text (.txt). The only aspect we have to pay attention to is that the first data row contains the variable names in text format and otherwise only numerical values (no text or special characters; also no numerical values in scientific format, e.g., 10E-7). The data we use with the reputation model can be downloaded either as a comma-separated value (.csv) or text (.txt) data set in the download section at https://www.pls-sem.com. Click on **Save Target As . . .** to save the data to a folder on our hard drive or cloud drive. Next, run the SmartPLS software. When we use the **New project** option in the toolbar, a new project will be created. Then, we can use the **Import data file** option in the newly created project. Alternatively, we can right-click on the project and use the **Import data file** option. With this data, as explained in Chapter 2 of the book by Hair, Hult, Ringle, and Sarstedt (2022), we can create the PLS path model as shown in Exhibit 1.11.

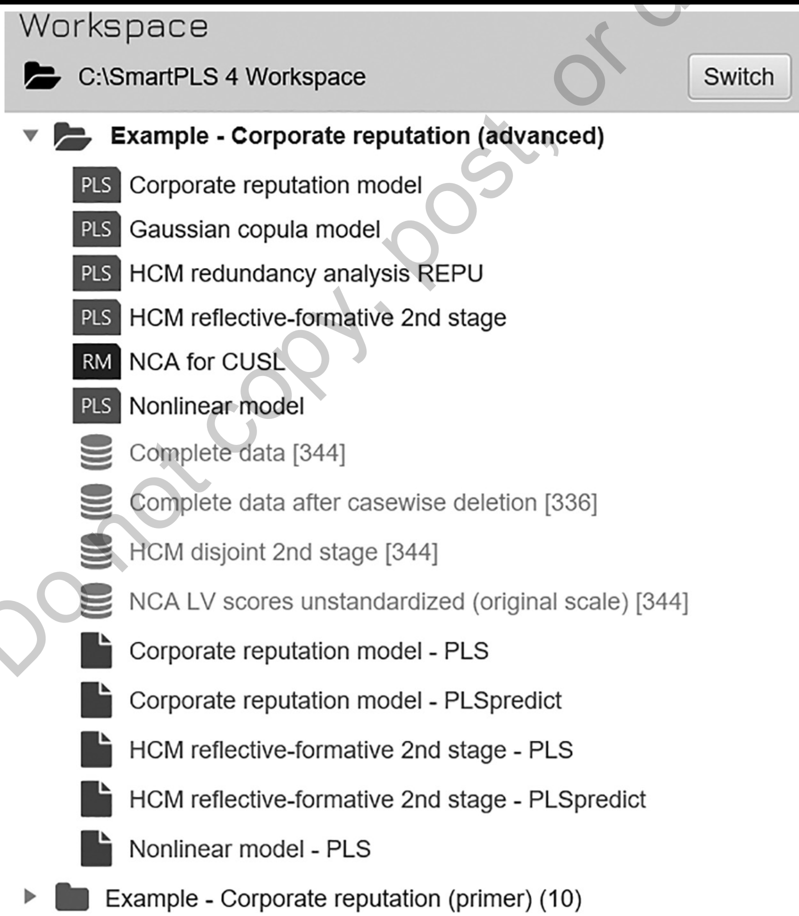
**EXHIBIT 1.11 ■ Corporate Reputation Model in SmartPLS**



Alternatively, after starting the SmartPLS software, various sample projects appear in the main window under **Sample Projects**, which can be installed directly. When clicking on the **Install** link next to **Example – corporate reputation (advanced)**, the advanced corporate reputation project will appear in the **Workspace** window (Exhibit 1.12), located at the left of the screen. Next, double-click on the **Corporate reputation model**. Then the PLS path model as shown in Exhibit 1.11 will appear in the SmartPLS modeling window.

Following the systematic procedure for applying PLS-SEM presented in Hair, Hult, Ringle, and Sarstedt (2022), the next steps entail the evaluation of the reflectively and formatively specified measurement models, followed by an assessment of the structural model. Readers are advised to consult the *Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (Hair, Hult, Ringle, & Sarstedt, 2022) for a detailed discussion and illustration of these analysis steps. The case study illustrations

**EXHIBIT 1.12 ■ The SmartPLS Workspace**



in the following chapters will depart from here, assuming that the quality of the original model's measurement and structural models have been established.

## SUMMARY

- **Understand the origins and evolution of PLS-SEM.** The precursors to PLS-SEM were two methods (i.e., principle component regression and PLS-R) that used least squares estimation to develop solutions for single and multicomponent models. Further development of these procedures by Herman Wold led to the NIPALS algorithm and a revised generalized version of the PLS algorithm that focused on finding constructs. In the 1980s Herman Wold proposed his "soft modeling basic design" underlying PLS-SEM as an alternative to CB-SEM. The latter method has been labeled as "hard" modeling due to its more stringent assumptions in terms of data distribution and sample size compared to PLS-SEM. While both approaches were developed at about the same time, CB-SEM became much more widely applied because of its early availability through the LISREL software in the late 1970s. It was not until the debut of Wynne Chin's PLS-Graph software in the mid-1990s that PLS-SEM applications began increasing. With the release of SmartPLS 2 in 2005, PLS-SEM's usage grew exponentially.
- **Comprehend model specification in a PLS-SEM framework.** The model specification in PLS-SEM involves two sub-models—the structural model and the measurement model. Whereas the structural model specifies the relationships between the constructs, the measurement models express how to measure the construct utilizing a set of indicators. Measurement models can be specified reflectively, using effect (i.e., reflective) indicators, or formatively, using causal or composite indicators. Whereas constructs measured with causal indicators have an error term, this is not the case with composite indicators, which define the construct in full. Traditionally, composite indicators have been viewed as a means to combine several variables to represent some new entity, whose meaning is defined by the choice of indicators. However, more recent research argues that composite indicators can be used to measure any type of property to which the focal concept refers, including attitudes, perceptions, and behavioral intentions.
- **Describe the PLS-SEM algorithm's basic functioning principles.** The PLS-SEM algorithm uses the empirical data for the indicators and iteratively determines the construct scores, the path coefficients, indicator loadings and weights, and further statistics, such as $R^2$ values and measures of the model's out-of-sample predictive power. After initialization, the algorithm estimates structural and measurement model parameters seperately, holding

the other model elements constant in each iteration. The algorithm's goal is to estimate parameters so residuals in the structural and measurement models are minimized. The results are typically standardized, meaning that, for example, the sizes of the path coefficients can be directly compared with each other even when the model estimation draws on differently scaled indicators. Recent extensions of the original PLS-SEM algorithm facilitate including sampling weights or estimating solutions comparable to common factor models.

- **Understand PLS-SEM's key characteristics vis-à-vis CB-SEM.** A crucial conceptual difference between PLS-SEM and CB-SEM relates to the way each method treats the constructs included in the model. CB-SEM considers the constructs as common factors, whereas PLS-SEM follows a composite model perspective using weighted composites of indicator variables to calculate scores that represent the constructs. Estimating a common factor model using PLS-SEM produces a bias, but the bias produced by CB-SEM when estimating composite models is much more substantial. In model estimation, PLS-SEM follows a causal-predictive paradigm in that the method seeks to maximize (in-sample) prediction of a specified model developed based on theory and logic. Because of the way PLS-SEM estimates the model parameters, the method is not constrained by identification issues, even if the model becomes complex and is being estimated with little data—a situation that typically restricts CB-SEM's use.

## REVIEW QUESTIONS

1. Who developed the PLS-SEM algorithm and what was the intention behind its development?
2. What is the difference between common factor models and composite models?
3. What is the difference between reflective and formative measurement models?
4. How does the PLS-SEM algorithm work?
5. What are the key characteristics that distinguish PLS-SEM from CB-SEM?

## CRITICAL THINKING QUESTIONS

1. Under what condition is PLS-SEM the preferred method over CB-SEM for prediction, and why?
2. Please comment on the following statement: "Indicators in formative measurement models are error-free."
3. What is the difference between causal and composite indicators?

4. Are common factor models and reflective measurement models the same?

5. Why should researchers test the predictive power of their PLS path models?

## KEY TERMS

Artifacts
Causal indicators
Common factor models
Common factor-based SEM
Common variance
Composite indicators
Composite-based SEM
Consistent PLS-SEM (PLSc-SEM)
Correlation weights
Covariance-based structural equation modeling (CB-SEM)
Effect indicators
Endogenous constructs
Error terms
Exogenous constructs
Explaining and predicting (EP) theories
Explanatory power
Factor (score) indeterminacy
Factor weighting scheme
Formative measurement model
In-sample predictive power
Inner model
Measurement models
Metrological uncertainty
Mode A
Mode B
Outer models
Partial least squares path modeling (PLS-SEM)
Partial least squares regression (PLS-R)
Path weighting scheme
PLS-SEM algorithm
PLS-SEM bias
Principal components regression
Reflective indicators
Reflective measurement model
Regression weights
Reliability coefficient $\rho_A$
Sampling weights
Single-item measurement
Structural model
Sum scores
Weighted PLS-SEM (WPLS)

## SUGGESTED READINGS


Bollen, K. A. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, *35*(2), 359–372.

Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, *22*(3), 581–596.

Chin, W. W., Cheah, J-H., Liu, Y., Ting, H., Lim, X.-J., & Cham, T. H. (2020). Demystifying the role of causal-predictive modeling using partial least squares structural equation modeling in information systems research. *Industrial Management & Data Systems*, *120*(12), 2161–2209.

Cook, D. R., & Forzani, L. (2023). On the role of partial least squares in path analysis for the social sciences. *Journal of Business Research*, *167*, 114132.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science*, *45*(5), 616–632.

Hair, J. F., Hult T., Ringle C. M., Sarstedt, M., Danks, N., & Ray, S. (2022). *Partial least squares structural equation modeling (PLS-SEM) using R: A workbook*. Springer.

Hair, J. F., Sarstedt, M., & Ringle, C. M. (2019). Rethinking some of the rethinking of partial least squares. *European Journal of Marketing*, *53*(4), 566–584.

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., Becker, J.-M., & Sarstedt, M. (2019). Factor indeterminacy as metrological uncertainty: Implications for advancing psychological measurement. *Multivariate Behavioral Research*, *54*(3), 429–443.

Rigdon, E. E., & Sarstedt, M. (2022). Accounting for uncertainty in the measurement of unobservable marketing phenomena. In H. Baumgartner & B. Weijters (Eds.), *Review of Marketing Research*, vol. 19 (pp. 53–73). Emerald.

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

Ringle, C. M., Sarstedt, M., Sinkovics, N., & Sinkovics, R. R. (2023). A perspective on using partial least squares structural equation modelling in data articles. *Data in Brief*, *48*, 109074.

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

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2023). PLS-SEM: Indeed a silver bullet–retrospective observations and recent advances. *Journal of Marketing Theory & Practice*, *31*(3), 261–275.

Schuberth, F., Müller, T., & Henseler, J. (2021). Which equations? An inquiry into the equations in partial least squares structural equation modeling. In I. Kemény & Z. Kun (Eds.), *New perspectives in serving customers, patients, and organizations: A Festschrift for Judit Simon* (pp. 96–115). Corvinus University of Budapest.

Yuan, K.-H., & Fang, Y. (2022). Which method delivers greater signal-to-noise ratio: Structural equation modelling or regression analysis with weighted composites? *British Journal of Mathematical and Statistical Psychology*. Advance online publication.