CHAPTER 1

An Overview of Recent and Emerging Developments in PLS-SEM

LEARNING OUTCOMES

1. Understand the origins and evolution of PLS-SEM.
2. Comprehend the principles of and recent developments in measurement.
3. Get to know the essential differences between PLS-SEM and CB-SEM and understand when to use each method.

CHAPTER PREVIEW

In recent years many developments have occurred in partial least squares structural equation modeling (PLS-SEM). Among the most prominent was the latest release of the SmartPLS software (Ringle, Wende, & Becker, 2015). User-friendly software makes social sciences scholars much more efficient and effective in reaching their research and publication goals. This is especially true when the software includes options to complete the most up-to-date analyses possible with a particular statistical method.

In this chapter, we first provide an overview of the origins and evolution of PLS-SEM to establish a foundation for better understanding why the method was slow to be adopted but has been increasingly applied in recent years across many social sciences disciplines, particularly the various fields of business administration. We then summarize the software that facilitates easy application of this rapidly emerging technique and briefly highlight recent methodological developments in PLS-SEM. To understand how PLS-SEM differs from covariance-based SEM (CB-SEM), we then discuss different approaches to measure conceptual variables and highlight which method is more suitable for selected model types.

ORIGINS AND EVOLUTION OF PARTIAL LEAST SQUARES STRUCTURAL EQUATION MODELING

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 for 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 revised generalized version of the **PLS algorithm** focused on establishing and including latent variables in path models (Lohmöller, 1989; Wold, 1980, 1982, 1985).

Several PLS procedures evolved from Wold's generalized least squares algorithm (Mateos-Aparicio, 2011). One procedure is **principal component regression**, which performs a principal component analysis on the independent variables and in which the principal components are used as predictive/explanatory variables for the dependent variable. But principal component regression focuses on reducing the dimensionality of the independent variables without taking into account the relationship between the independent and dependent variables.

Another procedure is **partial least squares regression (PLS-R)**, which was originally designed to reduce the problem of multicollinearity in regression models. PLS-R is an approach that focuses on dimension reduction of the independent variables in a regression model, with the objective of removing multicollinearity from the predictor variables, but doing so in a manner that 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, reduces the dimensionality of the independent variables, and takes into consideration the relationship between the independent and dependent variables, thus maximizing the explanation of the variance in the dependent variable. As a result, PLS-R allows researchers to estimate models with many more independent variables than observations in the data set (Valencia & Diaz-Llanos, 2003).

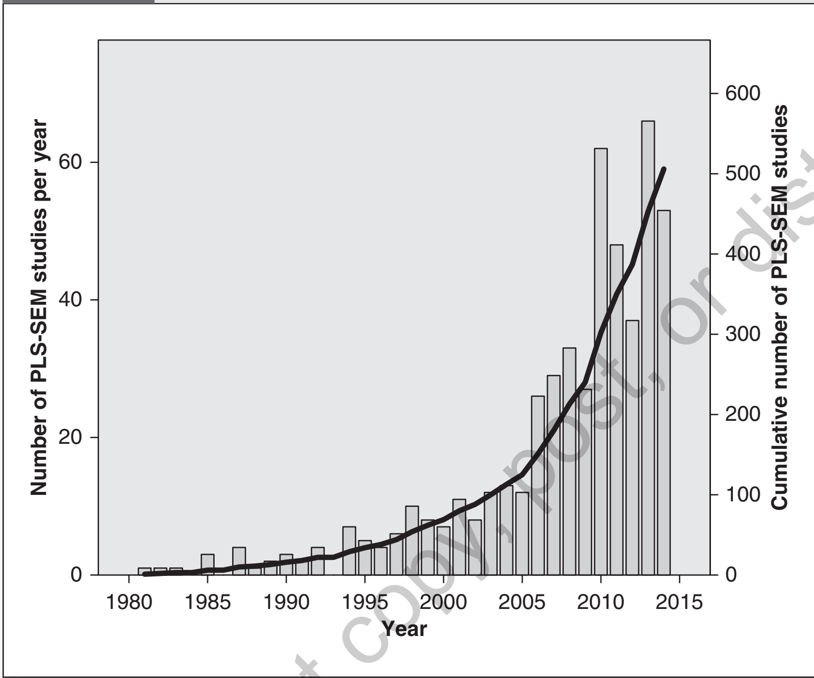
Interestingly, PLS-R was not developed by Herman Wold but by his son, Svante Wold, in the early 1980s. Svante Wold was working in the field of analytical chemistry that is known today as chemometrics—the application of statistical methods to chemical data. Together with Harald Martens, he adapted NIPALS to analyze chemical data, and in addition to solving 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.

A third procedure that emerged from Wold's generalized PLS algorithm was **partial least squares path modeling**, which later became known as **PLS-SEM** (Hair, Ringle, & Sarstedt, 2011). PLS-SEM determines the parameters of a set of equations in a structural model by combining principal component analysis to assess the measurement models with path analysis to estimate the relationships between latent variables. Wold (1982) proposed his "soft model basic design" underlying PLS-SEM as an alternative to Jöreskog's (1973) **covariance-based SEM (CB-SEM)**. CB-SEM is sometimes referred to as **factor-based SEM**, and has been labeled as hard modeling because of its much more rigorous assumptions in terms of data distribution and sample size. Importantly, "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 approaches were developed about the same time, CB-SEM became much more widely applied because of its early availability through the LISREL software since the late 1970s. In contrast, the first software for PLS-SEM was LVPLS, which appeared in the mid-1980s (Lohmöller, 1984) but was not very user-friendly. It was not until Chin's (1994) PLS-Graph software in the mid-1990s that PLS-SEM began being more widely applied. With the release of SmartPLS 2 in 2005 (Ringle, Wende, & Will, 2005), PLS-SEM applications grew exponentially.

Exhibit 1.1 summarizes the application of PLS-SEM in the top journals in marketing (Hair, Sarstedt, Ringle, & Mena, 2012) and strategic management (Hair, Sarstedt, Pieper, & Ringle, 2012), as well as *MIS Quarterly*, the flagship journal in management information systems research (Ringle, Sarstedt, & Straub, 2012). However, PLS-SEM use extends to many other fields as diverse as supply chain management, operations management, and family business research (e.g., Do Valle & Assaker, 2016; Hair, Hollingsworth, Randolph, & Chong, in press; Henseler, Ringle, & Sarstedt, 2012; Kaufmann & Gaeckler, 2015; L. Lee, Petter, Fayard, & Robinson, 2011; Nitzl, 2016; Peng & Lai, 2012; Richter, Sinkovics, Ringle, & Schlägel, 2016; Sarstedt, Ringle, Smith, Reams, & Hair, 2014). Also, several journals such as *Journal of Marketing Theory and Practice* (Hair et al., 2011), *Long Range Planning* (Hair, Ringle, & Sarstedt, 2013; Robins, 2012, 2014), *Journal of Business Research* (Carrión, Henseler, Ringle, & Roldán, 2016), and *International Marketing Review* (Richter et al., 2016) published special issues on the PLS-SEM method. Besides its popularity in business research, the use of PLS-SEM, as published journal articles reveal, has recently expanded into other research areas such as biology (e.g., Kansky, Kidd, & Knight, 2016), engineering (e.g., Kwofie, Adinyira, & Fugar, 2015; Liu, Zhao, & Yan, 2016; Minkman, Rutten, & van der Sanden, 2016), medicine (e.g., Hsu, Chang, & Lai, 2016; Pai, 2016; Pedrosa, Rodrigues, Oliveira, & Alexandre, 2016), and psychology (e.g., Jisha, & Thomas, 2016; Willaby, Costa, Burns, MacCann, & Roberts, 2015).

The release of version 3 of the most widely applied PLS-SEM software, SmartPLS, includes many new features. First, obtaining a basic solution from version 2 (Ringle et al., 2005) of the software was straightforward, and continues to be true for version 3. But quite a few diagnostics for assessing PLS-SEM solutions when using version 2 had to be calculated manually or completed with another software package, such as Excel, SPSS, or Statistica. SmartPLS 3 automatically performs these functions, including f^2 effect size, assessment of multicollinearity, and so forth. More important, however, is the addition of options such as confirmatory tetrad analysis (Gudergan, Ringle, Wende, & Will, 2008), the new heterotrait-monotrait ratio of correlations (HTMT) criterion to test discriminant validity (Henseler, Ringle, & Sarstedt, 2015), prediction-oriented segmentation (Becker, Rai, Ringle, & Völckner, 2013), moderator analysis (Henseler & Chin, 2010), different types of multigroup analysis (Sarstedt, Henseler, &

Exhibit 1.1 **Number of PLS-SEM Studies in Management, Marketing, and *MIS Quarterly***



Note: PLS-SEM studies published in *MIS Quarterly* were only considered from 1992 on.

Ringle, 2011), and invariance testing by means of the measurement invariance of composite models approach (Henseler, Ringle, & Sarstedt, 2016). The benefits of having these options readily at hand are tremendous, since these types of analyses are increasingly being requested by journal editors and reviewers. These developments are very important for social sciences researchers because many of them are not possible with CB-SEM due to its limiting assumptions, while the others have not been programmed to facilitate easy application.

MEASUREMENT

Conceptual Variables and Proxies

Whether researchers follow a deductive or inductive research approach, at some point—in their search to better understand and explain theory—they deal with theoretical models and conceptual variables. A **theoretical model** reflects a set of structural relationships, usually based on a set of equations connecting conceptual variables that formalize a theory and visually represent the relationships (Bollen, 2002). As elements of theoretical models, **conceptual variables** represent broad ideas or thoughts about abstract concepts that researchers establish and propose to measure in their research (e.g., customer satisfaction).

Constructs represent conceptual variables in statistical models such as in a structural equation model. They are intended to enable empirical testing of hypotheses regarding the conceptual variables (Rigdon, 2012) and are conceptually defined in terms of attribute and object (e.g., MacKenzie, Podsakoff, & Podsakoff, 2011). The **attribute** defines the general type of property to which the concept refers, such as an attitude (e.g., attitude toward an advertisement), a perception (e.g., perceived ease of use of technology), or a behavioral intention (e.g., purchase intention). The **focal object** is the entity to which the property is applied. For example, the focus of interest could be a customer's satisfaction with the products, satisfaction with the services, and satisfaction with the prices. In these examples, satisfaction constitutes the attribute, whereas products, services, and prices represent the focal objects.

Establishing a construct definition also includes determination of the dimensionality that describes the conceptual variable, with each dimension representing a different aspect of the conceptual variable. For example, the concept commitment can be viewed as encompassing three different dimensions (Meyer & Allen, 1991), each defined and measured differently: affective commitment (i.e., emotional attachment), continuance commitment (i.e., the gains versus losses of working at an organization), and normative commitment (i.e., based on feelings of obligation). Each of these dimensions requires a separate operational definition. A conceptual variable is not per se characterized as unidimensional or multidimensional, let alone two-, three-, or four-dimensional (Bollen, 2011). Rather it depends on the context-specific definition of the conceptual variable and the denotation that comes with it. The denotation can, in principle, be infinite since the same conceptual variable may represent different levels of theoretical abstraction across contexts (Diamantopoulos, 2005; Law & Wong, 1999). An operational definition is subject to the context within which a conceptual variable is examined, and the definition can change from one study to another. Consequently, an operational definition can differ in terms of dimensionality and the object of interest, depending on the context of the study. For example, a customer's satisfaction with the service can be broken down into more concrete subdimensions, such as satisfaction with the speed of service, the servicescape, and the staff. The latter dimension can be divided into more concrete subdimensions such as satisfaction with the friendliness, competence, and outer appearance of the service staff. Each of these aspects can, in principle, be further broken down into yet more concrete subdimensions.

As the construct definition clarifies how the abstract, conceptual variable relates to measurable, observable quantities, it also assists in understanding how, in structural equation models, conceptual variables are represented by **constructs** (also referred to as **latent variables**). Constructs are not directly observed but rather inferred (mathematically) from **manifest variables** that are observed (directly measured). Manifest variables are also referred to as **items** or **indicators**. The indicators correspond, for example, to questions in a survey capturing respondents' perceptions, attitudes, and behaviors, or they represent some other characteristic of the object under investigation (e.g., a customer's satisfaction with the product). Importantly, manifest variables do not need to be survey-based data but can also be secondary data (e.g., satisfaction ratings from a review website). When a conceptual variable is multidimensional, each dimension needs to be represented by a separate construct, operationalized with a specific set of manifest variables. For example, if a study considers all three dimensions of commitment, then there will be three separate latent variables in the structural equation model.

The construct definition explains how the conceptual variable should relate to manifest variables from a theoretical perspective. However, constructs do not represent conceptual variables perfectly as any concept and any operational definition has some degree of ambiguity associated with it. Furthermore, constructs stem from data and therefore share the data's idiosyncrasies (Cliff, 1983; MacCallum, Browne, & Cai, 2007; Rigdon, 2012), which further detaches them from the concepts they intend to represent. In this context, Michell (2013, p. 20) notes that constructs

> are contrived in a way that is detached from the actual structure of testing phenomena and held in place by an array of quantitative methods, such as factor analysis, which gratuitously presume quantitative structure rather than infer it from the relevant phenomena.

Against this background, Rigdon (2012, pp. 343–344) concludes that constructs should rather be viewed as "something created from the empirical data which is intended to enable empirical testing of propositions regarding the concept." That is, all measures of conceptual variables are approximations of or proxies for conceptual variables, independent from how they were derived (e.g., Wickens, 1972). Thus, 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 latent variables (Rigdon, 2012). This assessment is in line with the proliferation of all sorts of instruments that claim to measure essentially the same construct, albeit often with little chance to convert one instrument's measures into any other instrument's measures (Salzberger, Sarstedt, & Diamantopoulos, 2016). For example, business research and practice has brought forward 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).

Measurement Models

Following the construct definition, the next step is to conceptualize a **measurement model**, which expresses how to measure the construct by means of a set of indicators. Generally, there are two broad ways to conceptualize measurement models. 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 with relationships going from the construct to its indicators (Exhibit 1.2). The relationship between an observed and an unobserved variable is usually modeled as expressed in the following equation:

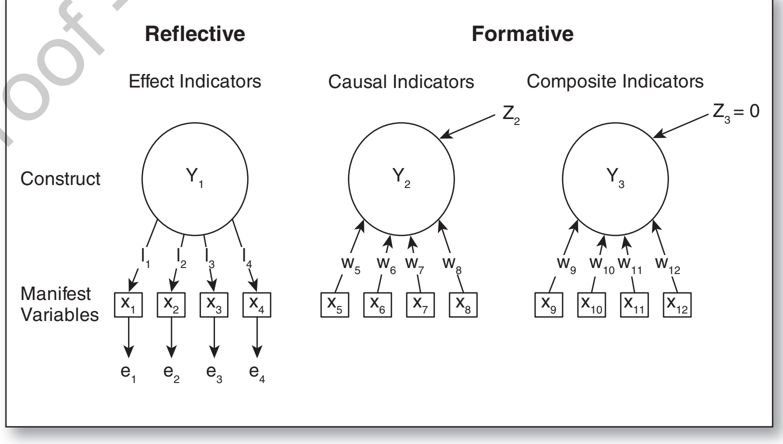
$$x = l \cdot Y + e,$$

where x is the observed indicator variable, Y is the latent variable, the loading l is a regression coefficient quantifying the strength of the relationship between each indicator x and its construct Y; e represents the random measurement error.

Exhibit 1.2 illustrates a **reflective measurement model** for a latent variable Y_1, measured with four indicators x_1, x_2, x_3 and x_4. Here, the l_1 to l_4 relationships result from four simple regressions, as depicted by the previous equation, with each indicator x_1 to x_4 as dependent variable and construct Y_1 as independent variable. **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 dictates 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

Exhibit 1.2 Measurement Model Types



construct by means of 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 latent variable (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.

Despite these clear conceptual differences, deciding whether to specify measurement models reflectively or formatively is not clear-cut in practice, as conceptual variables do not inherently follow a reflective or formative measurement logic. Rather, the researcher has the flexibility to define how such proxies are to be derived in a measurement model based on the construct definition the researcher specifies. Consider, for example, the concept of perceived switching costs. Jones, Mothersbaugh, and Beatty (2000, p. 262) define perceived switching costs as "consumer perceptions of the time, money, and effort associated with changing service providers." Their measurement approach in the context of banking services draws on three items, which constitute reflections or consequences of perceived switching costs ("In general it would be a hassle changing banks," "It would take a lot of time and effort changing banks," and "For me, the costs in time, money, and effort to switch banks are high"). Hence, the authors implicitly assume that there is a concept of perceived switching costs, which can be manifested by querying a set of (e.g., three) items. Barroso and Picón (2012, p. 532), on the other hand, consider perceived switching costs as "a latent aggregate construct that is expressed as an algebraic composition of its different dimensions." These authors identify a set of six dimensions (benefit loss costs, personal relationship loss costs, economic risks costs, evaluation costs, setup costs, and monetary loss costs), which represent certain specific characteristics, each covering an independent part of the perceived switching costs concept. As such, Barroso and Picón's conceptualization of perceived switching costs represents a formative measurement model logic. Of course, the underlying items are empirically correlated, and perhaps causally related, but they are not actually exchangeable in the way the reflective measurement model conceptualization assumes they are (Rigdon, Preacher, et al., 2011). That is, their correlation is not because the construct of perceived switching costs is assumed to be their common cause.

Further contributing to the difficulties of deciding on the measurement perspective is the fact that there is not one type of formative measurement model—as suggested in the early works on formative measurement (e.g., Diamantopoulos & Winklhofer, 2001) and the use of formative measurement models in statistical analysis (e.g., Hair et al., 2011). Rather, two types of indicators exist in formative measurement models: causal indicators and composite indicators (Bollen, 2011; Bollen & Bauldry, 2011). As implied by their name, **causal indicators** are assumed to cause the underlying construct. Therefore, the indicators should have conceptual unity in that all the indicators correspond to the researcher's definition of the concept (Bollen & Diamantopoulos, 2016). Breadth of coverage of the domain is extremely important to ensure that the domain of content is adequately captured—omitting important indicators implies omitting a part of the conceptual variable that the construct represents. As causal indicators are expected to cover all aspects of the content domain (Bollen & Bauldry, 2011), constructs measured with causal indicators (construct Y_2 with indicators x_5 to x_8 in Exhibit 1.2) have an error term (z_2 in Exhibit 1.2). This error term captures all the other causes of the construct not included in the model (Diamantopoulos, 2006). Importantly, the indicators themselves are assumed to be error-free. The following equation represents a measurement model with causal indicators, where w_i indicates the contribution of x_i ($i = 1, \ldots, I$) to Y, and z is an error term associated with Y:

$$Y = \sum_{i=1}^{I} w_i \cdot x_i + z.$$

The other indicator type, **composite indicators**, closely resembles causal indicators except for a small detail. Similar to constructs measured with causal indicators, constructs measured with composite indicators (construct Y_3 with indicators x_9 to x_{12} in Exhibit 1.2) also have an error term but this term is set to zero ($z_3 = 0$ in Exhibit 1.2). This distinction has an essential implication for the characterization of formative measurement models (Henseler et al., 2014), since composite indicators operate as contributors to a construct rather than truly causing it (Bollen, 2011; Bollen & Bauldry, 2011). They form the composite representing the construct fully by means of linear combinations, thereby ensuring that the construct has no error term. The following equation illustrates a measurement model with composite indicators, where *Y* is a linear combination of indicators x_i, each with an indicator weight w_i (Rigdon, 2012):

$$Y = \sum_{i=1}^{I} w_i \cdot x_i.$$

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 (e.g., Bollen, 2011; Henseler, Hubona, & Ray, 2016). For example, a measurement model conceptualization of information search activities could be based on capturing the sum of the activities that customers engage in when seeking information from dealers, promotional materials, the Internet, and other sources. Another researcher might choose a different set of variables to form a measure of information search activities. Thus, the items ultimately determine the meaning of the construct, which implies that adding or omitting an indicator potentially alters the nature of the construct. Therefore, according to Bollen (2011), composite indicators need not share conceptual unity but may have a similar theme.

In practice, however, it remains largely unclear where to draw a line between items having conceptual unity and sharing a similar theme. In fact, researchers naturally adopt a construct definition of the concept and choose items in operationalizing measurement models that match their definition, regardless of whether the actual measurement conceptualization draws on reflective, causal, or composite indicators. That is, they treat the constructs in their studies as unitary entities, just like Barroso and Picón (2012) do when offering an in-depth literature review of the nature and dimensionality of the perceived switching costs concept, prior to deriving indicators in their operationalization of the construct's measurement model. As such, they follow best practice in measurement by developing a guiding conceptual framework and set of indicators, which imbues the construct's theoretical meaning.

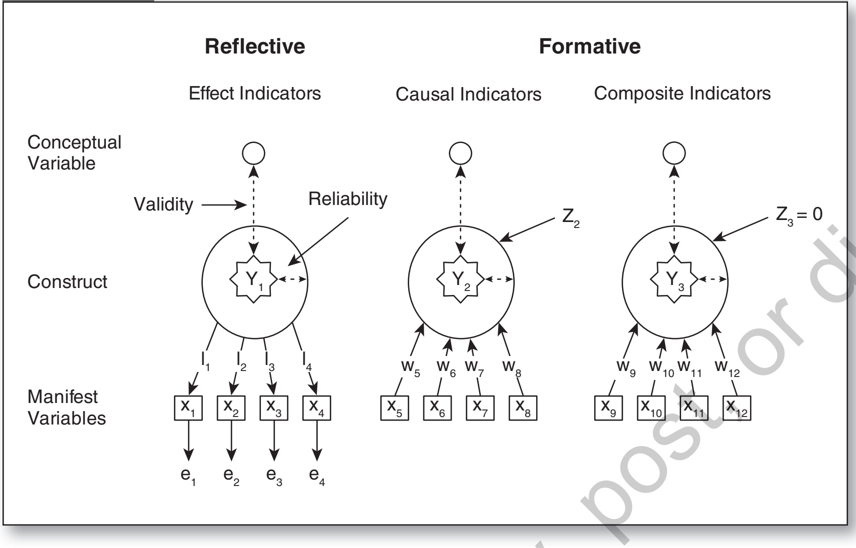
Assuming that researchers use measures of composite indicators merely as convenient summaries of the data—as some researchers do—implies that the common practice of aggregating items as composites to represent constructs, even though commonly done in practically all non-SEM studies in all fields of research (e.g., when using composites as input for a regression analysis), is without any theoretical justification and undermines the fundamentals of appropriate measurement. However, the very same measures have mostly been carefully developed and tested following common measurement model evaluation guidelines—as extensively documented in standard measurement scale handbooks (e.g., Bearden, Netemeyer, & Haws, 2011; Bruner, James, & Hensel, 2001). Thus, the very activity of forming composites from validated measurement scales interweaves composite and causal indicators. Therefore, calling for an abandonment of PLS-SEM because it uses composites as representations for conceptual variables (Guide & Ketokivi, 2015) implies an abandonment of practically all empirical research involving latent variables, including studies published in operations management, marketing, strategy, and other social sciences.

Composite indicators not only offer a convenient way to summarize the data but can also be used to measure any type of property to which the focal concept refers, including attitudes, perceptions, and behavioral intentions. However, as with any type of measurement conceptualization, researchers need to offer a clear definition and define items that closely match this definition—that is, they must share conceptual unity. Alternatively, measurement models with composite indicators can be interpreted as a prescription for dimension reduction, where the aim is to condense the measures so they adequately cover a conceptual variable's salient features (Dijkstra & Henseler, 2011). For example, a researcher may be interested in measuring the salient aspects of perceived switching costs by means of three (composite) indicators, which cover aspects particularly relevant to the study at hand (e.g., evaluation costs, setup costs, monetary loss costs).

Exhibit 1.3 extends the previous display of reflective and formative measurement model conceptualizations to include the conceptual variables they seek to represent (Rigdon, 2012). Conceptual variables are the small circles at the top of the diagrams. The proxy estimates are shown as the smaller eight-point star-like symbols in the middle of the larger circles. The vertical double-headed arrows suggest the extent to which the proxies are valid measures—the shorter these vertical arrows, the more valid the proxies.

The position of the proxy within the larger circle is an indication of the reliability of the construct measures, and the middle represents

Exhibit 1.3 Measurement Model Types and Conceptual Variables



perfect reliability. Note that we have to consider different types of reliability, depending on the model type. Internal consistency reliability statistics such as Cronbach's alpha apply to reflective measurement models. But these statistics are not appropriate when evaluating formative measurement models because causal and composite indicators do not necessarily have to correlate. In formative measurement models, test-retest reliability is the only means of testing the measures' reliability. However, in light of the manifold problems of assessing test-retest reliability (e.g., Campbell & Stanley, 1966), most researchers routinely disregard reliability assessment when estimating formative measurement models.

MODEL ESTIMATION

The previous sections describe different measurement model types and their indicators. The next question is how to estimate these models and their relationships with others in the structural equation model. Researchers typically use two approaches to estimate structural equation models: CB-SEM (Bollen, 1989; Diamantopoulos, 1994; Jöreskog, 1978) and PLS-SEM (Hair, Hult, Ringle, &

Sarstedt, 2017; Lohmöller, 1989; Wold, 1982). While both methods estimate the relationships among constructs and indicators, they differ fundamentally in their approaches in doing so (Jöreskog & Wold, 1982).

CB-SEM initially divides the variance of each indicator into two parts: (1) the **common variance**, which is estimated from the variance shared with other indicators in the measurement model of a construct, and (2) the **unique variance**, which consists of both **specific variance** and **error variance**. The specific variance is assumed to be systematic and reliable, while the error variance is assumed to be random and unreliable (i.e., measurement, sampling, and specification error). CB-SEM initially calculates the covariances of a set of variables (common variance), and only that variance is included in any solutions derived. CB-SEM, therefore, follows a **common factor model approach** in the estimation of the construct measures, which hypothesizes that the variance of a set of indicators can be perfectly explained by the existence of one unobserved variable (the common factor) and individual random error (Spearman, 1927; Thurstone, 1947). The common factor model estimation approach fully conforms to the measurement philosophy underlying reflective measurement models.

In principle, CB-SEM can also accommodate formative measurement models with causal indicators, but to ensure model identification, researchers must follow model specification rules that require certain constraints on the model (Bollen & Davis, 2009; Diamantopoulos & Riefler, 2011). MacCallum and Browne (1993), among others, have addressed these CB-SEM specification issues in detail. However, as Hair, Sarstedt, Ringle, and Mena (2012, p. 420) note, "these constraints often contradict theoretical considerations, and the question arises whether model design should guide theory or vice versa." As an alternative, CB-SEM scholars have proposed the **multiple indicators and multiple causes (MIMIC) model**, which includes both formative and reflective indicators (e.g., Diamantopoulos & Riefler, 2011; Diamantopoulos, Riefler, & Roth, 2008). While MIMIC models enable researchers to deal with the identification problem, they do not overcome the problem that formative measurement models with causal indicators invariably underrepresent the variance in the construct, since correlated indicators are required by the CB-SEM common factor model to produce a valid proxy and thereby adequately represent a conceptual variable. As Lee and Cadogan (2013,

p. 243) note, "researchers should not be misled into thinking that achieving statistical identification allows one to obtain information about the variance of a formative latent variable." In summary, application of the MIMIC approach to include constructs measured with causal indicators in CB-SEM invariably introduces a bias in the proxies and therefore in the SEM solutions. Clearly, at best CB-SEM only allows for approximating formative measurement models with causal indicators.

Similarly, CB-SEM can accommodate formative measurement models with composite indicators (Bollen & Diamantopoulos, 2017). Since constructs measured with composite indicators are defined by having zero variances, the identification of the construct's error variance is not an issue. However, problems arise with regard to the identification of all paths leading to as well as flowing out from the construct. Grace and Bollen (2007, p. 206) note that this problem can be solved by specifying a single incoming or outgoing path relationship to 1.0 (e.g., one relationship from the indicators to the composite per measurement model). While such specifications overcome parameter identification issues, they severely limit the interpretability of the structural model estimates with regard to the significance and magnitude of the fixed paths. Furthermore, constructs with composite indicators can only be included as exogenous latent variables while at the same time requiring at least one relationship to a reflectively measured endogenous latent variable. Both of these issues limit their usefulness in practice.

Different from CB-SEM, PLS-SEM does not divide the variance into common and unique variance. More precisely, the objective of PLS-SEM is to account for the total variance in the observed indicators rather than to explain only the correlations between the indicators (e.g., Chin, 1998; Tenenhaus, Esposito Vinzi, Chatelin, & Lauro, 2005). The differing treatment of the variances in PLS-SEM compared to CB-SEM corresponds to the distinction between principal component analysis and common factor analysis. Whereas principal components analysis assesses total variance and estimates factors (components) as linear combinations of indicators, common factor analysis assesses shared (common) variance only (Kline, 2015). The logic of the PLS-SEM approach is that in estimating the model relationships, all of the variance (common, unique, and error) that the exogenous variables have in common with the endogenous variables should be included. The underlying notion of PLS-SEM is that the indicators can be (linearly) combined to form composite variables that are comprehensive representations of the latent variables, and that these linear combinations are valid proxies of the conceptual variables under investigation. Therefore, PLS-SEM follows a **composite model approach** in the estimation of the construct measures. PLS-SEM's designation as composite-based only refers to the method's way to represent constructs that approximate the conceptual variables in a model. But while PLS-SEM draws on composites whose use has traditionally been considered to be consistent with formative measurement models and not reflective measurement models, the fact is the method readily accommodates both measurement model types (e.g., Roldán & Sánchez-Franco, 2012). In formative measurement models, however, PLS-SEM treats all indicators as composite indicators. That is, the method does not allow for the explicit modeling of a construct's error term measured with causal indicators (i.e., the error term z_3 in Exhibits 1.2 and 1.3 is zero). As a consequence and analogous to CB-SEM, PLS-SEM only allows for approximating formative measurement models with causal indicators.

Finally, to estimate the model parameters, PLS-SEM uses two modes, which relate to the way the method estimates the indicators weights that represent each indicator's contribution to the composite. Mode A corresponds to correlation weights derived from bivariate correlations between each indicator and the construct; Mode B corresponds to regression weights, the standard in ordinary least squares regression analysis. Regression weights take into account not only the correlation between each indicator and the construct but also the correlations between the indicators. No matter which mode for estimating the indicator weights is used, the resulting latent variable is always modeled as a composite (Henseler, Hubona, & Rai, 2016). That is, since all multi-item measures are converted into weighted components—even in Mode A—PLS-SEM computes components by means of linear combinations of indicators.

PLS-SEM by default uses Mode A for reflectively measured constructs and Mode B for formatively measured constructs. Recent research, however, suggests that selecting the appropriate weighting mode requires a more thoughtful approach. Specifically, Becker, Rai, and Rigdon (2013) show that for formatively measured constructs, Mode A estimation yields better out-of-sample prediction for sample sizes larger than 100 and when the R^2 value is moderate to large (i.e., $R^2 > 0.30$). For large sample sizes and large R^2 values, Mode A and Mode B perform equally well in terms of out-of-sample prediction. In terms of parameter accuracy in the structural model, Mode A performs best when sample size or R^2 values are small to medium. For larger sample sizes or R^2 values, Mode A and Mode B estimations do not differ in terms of parameter accuracy.

The option of using different estimation modes per construct allows tailoring the model estimation to fit the study's objective. Similarly, item weights of certain construct measures can also be defined a priori (Howell, Breivik, & Wilcox, 2013), thereby maintaining temporal constancy in the relative importance of each item in the formation of the composite. While such a step facilitates a consistent interpretation in repeated assessments of the same model, fixed weights equalize any differences in each item's importance for forming the composite in a specific context. However, understanding individual item weights offers important insights in the specific context of the analysis. When measuring, for example, service satisfaction, weights denote which individual items are of particular importance in the shaping of service satisfaction. Although individual item weights likely differ across contexts, as all items share conceptual unity, the empirical meaning of the construct (e.g., Howell, Breivik, & Wilcox, 2007) will differ only within the realm of the construct definition.

PLS-SEM OR CB-SEM?

Researchers have long acknowledged the differences between PLS-SEM and CB-SEM, highlighting situations that favor the use of one method over the other (e.g., Jöreskog & Wold, 1982). Similarly, an abundance of studies have empirically compared the differences in model estimates produced by CB-SEM versus PLS-SEM. These comparisons univocally focused on PLS-SEM's capabilities to obtain the same results as CB-SEM. However, PLS-SEM in its original form (Wold, 1982; Lohmöller, 1989) was not created to mimic CB-SEM but follows a different philosophy of measurement and aim of the analysis. In the following sections, we discuss these two key criteria that primarily distinguish PLS-SEM from CB-SEM. Hair et al. (2017) offer a more detailed discussion of model and measurement characteristics relevant for choosing between PLS-SEM and

CB-SEM. Finally, Rigdon (2016) and Sarstedt, Hair, and Ringle (2017) discuss reasons for using PLS-SEM and flawed arguments in favor of the method.

Philosophy of Measurement

A crucial conceptual difference between PLS-SEM and CB-SEM relates to the way each method treats the latent variables included in the model. As CB-SEM considers only the indicators' common variance, the method treats the constructs as common factors. PLS on the other hand uses the indicators' total variance in that the method generates linear combinations of indicators to represent the constructs, thereby constituting a composite model approach to SEM.

Several simulation studies have shown that PLS-SEM measurement model estimates (loadings) are larger and structural model estimates (path coefficients) are smaller relative to those obtained by CB-SEM (e.g., Reinartz, Haenlein, & Henseler, 2009). However, these simulation studies univocally defined common factor populations and drew data from these populations (Marcoulides & Chin 2013). Therefore, they evaluated PLS-SEM on the basis of (common factor–based) populations that are inconsistent with the method's philosophy of measurement. Indeed, practically every statistical method will perform less well when the underlying model is misspecified, and PLS-SEM is no exception in this regard. In fact, CB-SEM is known to perform less well when the model being estimated is inconsistent with the population, such as when the data are generated from a composite model (Becker, Rai, & Rigdon, 2013; Henseler et al., 2014). Becker, Rai, and Rigdon (2013) defined composite-based populations, and showed that PLS-SEM parameter estimates are the same as those of CB-SEM—converging on those values as sample size increases. Thus, the term *consistency at large* when applied to PLS-SEM results as representing bias is a misnomer.

CB-SEM results based on the common factor model, including the path coefficients, inter-construct correlations, and indicator loadings, have been labeled as the "true score" (Bollen, 1989). But what makes them so—other than that they have been called that previously? What is a true score when measuring concepts such as attitudes, perceptions, or intentions? Researchers have long warned that the common factor model rarely holds in applied research (Schönemann & Wang, 1972). For example, among 72 articles published during 2012 in what Atinc, Simmering, and Kroll (2012) consider the four leading management journals that tested one or more common factor model(s), less than 10% contained a common factor model that did not have to be rejected. In light of these and similar results, neither model can be assumed to carry greater significance than the other with regard to the existence or nature of conceptual variables (Rigdon, 2016). Any construct operationalization—regardless of whether based on a common factor or composite models logic—comes with an abundance of ambiguities related to, for example, the construct definition (e.g., Diamantopoulos, 2005; Gilliam & Voss 2013), the item wordings, and the number of items necessary to capture the construct domain (e.g., DeVellis, 2011). The same holds for the measurement validation, which is highly context-specific, thereby capitalizing on the idiosyncrasies of the data at hand. In light of these ambiguities, it is more reasonable to view estimates produced by both CB-SEM and PLS-SEM as proxies for the concepts under research and nothing more (Rigdon, 2012).

Irrespective of these conceptual concerns, the differences that PLS-SEM produces when estimating common factor models is very small provided that the measurement models meet minimum recommended standards in terms of the number of indicators and indicator loadings, and the model is correctly specified. Specifically, when the measurement models have four or more indicators and indicator loadings meet recommended standards, there is very little bias in parameter estimates when estimating a common factor model with PLS-SEM, as shown by Reinartz et al. (2009).

Prior efforts to dramatize the differences between CB-SEM and PLS-SEM estimates (Rönkkö, McIntosh, Antonakis, & Edwards, 2016) in Reinartz et al.'s (2009) study focused on descriptive differences between population values and parameter estimates only, disregarding the simple concept of statistical inference. As Reinartz et al. note in their results description of all simulation conditions, "parameter estimates do not differ significantly from their theoretical values for either ML-based CB-SEM (p-values between 0.3963 and 0.5621) or PLS-SEM (p-values between 0.1906 and 0.3449)" (p. 338). Only when the model estimation draws on a high sample size ($N = 10{,}000$) and measurement models with many indicators with high loadings, statistically significant differences between 0.58% and 3.23% occurred. Most importantly, when estimating common factor models, the divergence between PLS-SEM and CB-SEM results is of little practical relevance for the vast majority of applications (e.g., Astrachan, Patel, & Wanzenried, 2014). In fact, recent research shows that the bias that CB-SEM produces when (incorrectly) using the method to estimate composite models is much higher than the bias that PLS-SEM produces when using the method to (incorrectly) estimate common factor models (Sarstedt, Hair, Ringle, Thiele, & Gudergan 2016). The same research shows that PLS-SEM's parameter bias is clearly lower when correctly using the method to estimate composite models than CB-SEM's bias when using the method for estimating common factor models, especially when sample sizes are small. To summarize, PLS-SEM introduces practically no bias when estimating data from a composite model population, regardless of whether the measurement models draw on composite indicators or effect indicators. Biases are somewhat higher for factor model populations, but low in absolute terms. Clearly, PLS-SEM is optimal for estimating composite models while simultaneously allowing approximating common factor models with effect indicators with practically no limitation (Sarstedt, Hair, et al., 2016).

The minimal magnitude of differences between composite-based PLS-SEM and common factor–based CB-SEM when the underlying data stems from a common factor model population calls into question the need for "corrections" of the PLS-SEM method when estimating factor models. Specifically, in an effort to align common factor and composite-based SEM methods, Dijkstra and Henseler (2015a, 2015b) recently introduced the **consistent PLS** (**PLSc**) approach (also see Bentler & Huang, 2014). PLSc follows a composite modeling logic but mimics a common factor model. To do so, the method first estimates the model parameters using the standard PLS-SEM algorithm and corrects these estimates for attenuation using the consistent reliability coefficient ρ_A. This correction applies only to reflective measurement models; formative measurement model estimates of composite indicators remain unchanged. That is, when estimating reflective measurement models, PLSc draws on the questionable premise that CB-SEM results are the benchmark against which PLS-SEM results should be evaluated. Several researchers have noted that PLS-SEM-based indicator loadings are upwardly biased, but that is in comparison to indicator loadings from CB-SEM. The divergence of PLS-SEM parameter estimates should not be considered a bias, but a different result based on an algorithm following a different philosophy of measurement.

Generally, it is difficult to argue why PLS-SEM users would want the method to produce the same results as CB-SEM, since that method is already widely recognized and accepted as an approach to obtain solutions to research designed around the common factor model. One potential reason might be that the model is underidentified or its estimation leads to convergence problems in CB-SEM. Alternatively, researchers may want to include common factors and composites in the same model and ensure that each type is estimated in accordance with its conceptualization.

Aim of the Analysis

Apart from the differences in the philosophy of measurement and the different treatment of latent variables, the nature of **latent variable scores** as produced by PLS-SEM and CB-SEM also has consequences for the methods' areas of application. In CB-SEM latent variable scores are not unique, which means that there is an infinite number of different sets of latent variable scores that will fit the model equally well. A crucial consequence of this **factor (score) indeterminacy** is that the correlations between a common factor and any variable outside the factor model are indeterminate (Guttman, 1955). That is, a correlation may be high or low, depending on which set of factor scores one chooses. As a result, this limitation makes CB-SEM extremely unsuitable for prediction (e.g., Dijkstra, 2014). In contrast, PLS-SEM always produces a single specific (i.e., determinate) score for each composite for each observation, once the weights are established. Using these proxies as input to obtain a solution, PLS-SEM applies ordinary least squares regression with the objective of maximizing the R^2 values of the endogenous constructs. PLS-SEM is therefore the preferred method when the aim of the analysis is prediction (Albers, 2010; Rigdon, 2014; Sarstedt, Ringle, Henseler, & Hair, 2014).

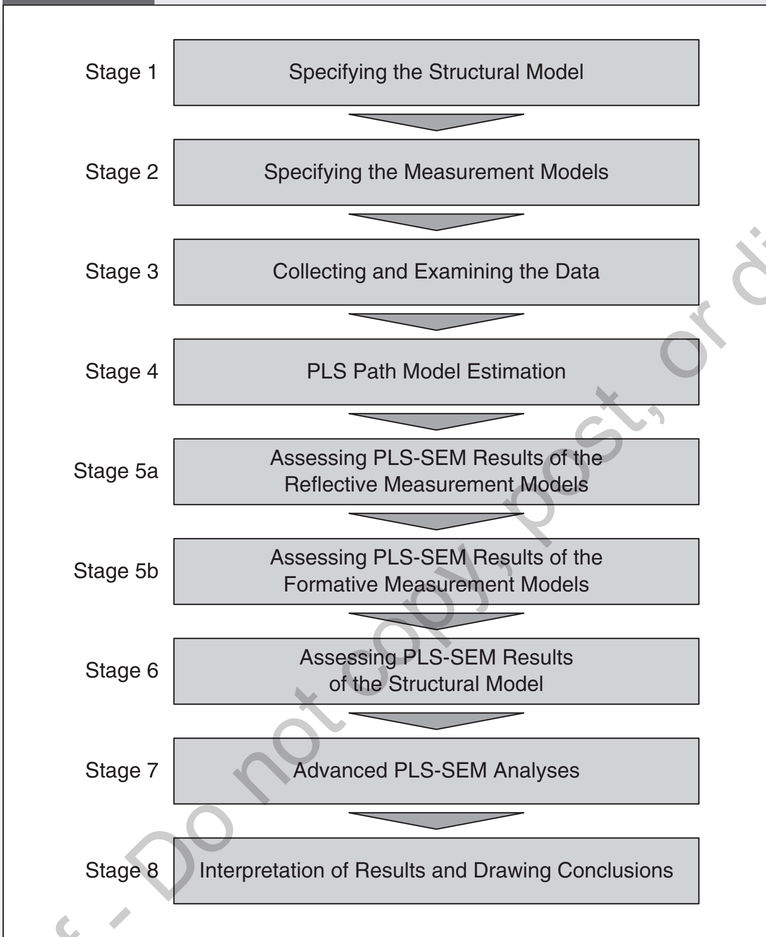
Researchers should be aware that the R^2 values in both types of SEM cannot be interpreted the same as with multiple regression models. The R^2 value in SEM is the variance explained in the construct scores, and not the variance explained in the indicators. Since the construct score represents only a portion of the variance in the indicators, the R^2 value in SEM is the amount of variance explained in the construct. The amount of variance included in the construct score is represented by the **average variance extracted** (**AVE**), which is the degree to which the construct explains the variance of its indicators. Thus the R^2 value in SEM is the variance explained in the AVE.

As an example of the R^2 value in SEM, consider the following. CB-SEM is based only on common variance. If you assume common variance representing the indicators of an endogenous construct is 60% of the total variance, and the AVE is 0.60, then the variance in the AVE is actually 36% of the original variance in the total variance (60% of 60% = 36%). Thus, if the CB-SEM R^2 value is moderately high, for example, 0.50, then the proportion of the total variance predicted in the endogenous construct is only 18% of the total variance in the indicators. In contrast, consider a similar example with PLS-SEM that is based on the total variance. If the AVE of an endogenous construct is 0.60, and the PLS-SEM R^2 value is 0.50, then the proportion of the total variance predicted in the endogenous construct is 30% of the total variance in the indicators of the endogenous construct, almost twice as much. PLS-SEM is based on total variance, and AVEs are always higher than in CB-SEM, so the R^2 value in PLS-SEM is more meaningful as a predictor of the variance included in the indicators of the endogenous constructs than is the R^2 value in CB-SEM. To summarize, when the research objective is prediction, the choice of methods is clear—it should be PLS-SEM.

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 3 software for such analyses. In this advanced book, we expand on Stage 7 of the systematic procedure for applying PLS-SEM (Exhibit 1.4) from the *Primer on Partial Least Squares Structural Equation Modeling,* 2nd edition *(PLS-SEM)* (Hair et al., 2017). The advanced PLS-SEM analyses will enable you to better understand and explain your results, and provide the types of analysis and diagnostic metrics editors and reviewers increasingly request. Exhibit 1.5 identifies the chapters and topics covered in this book.

Exhibit 1.4 A Systematic Procedure for Applying PLS-SEM



Chapter 2 offers an introduction to advanced modeling topics, starting with hierarchical component models. A hierarchical component model represents a more general construct, measured at a higher level of abstraction, while simultaneously including several subcomponents, which cover more concrete traits of the conceptual variable represented by this construct. With the increasing complexity of theories and cause-effect models in the social sciences, researchers have more often used these models in their PLS-SEM studies (e.g., Ringle et al., 2012). Similarly, there has also been increased interest in modeling nonlinear relationships between constructs. When the relationship

Exhibit 1.5 Chapters and Topic in This Book

Chapter	*Topics*
2	Advanced Modeling – Hierarchical component models – Nonlinear relationships
3	Advanced Model Assessment – Confirmatory tetrad analysis (CTA-PLS) – Importance-performance map analysis (IPMA)
4	Modeling Observed Heterogeneity – Measurement invariance assessment (MICOM) – Multigroup analysis
5	Modeling Unobserved Heterogeneity – Finite mixture PLS (FIMIX-PLS) – PLS prediction-oriented segmentation (PLS-POS)

between two constructs is nonlinear, the size of the effect between two constructs depends not only on the magnitude of change in the exogenous construct but also on its value. In the second part of Chapter 2, we introduce the principles of nonlinear modeling and describe how to run corresponding analyses in SmartPLS 3.

In Chapter 3, we discuss two types of advanced model assessment, starting with the confirmatory tetrad analysis (CTA-PLS), which allows empirically assessing whether data support a formative measurement model specification or a reflective specification. Next, we introduce the importance-performance map analysis (IPMA), which extends the standard structural model results reporting of path coefficient estimates by adding a dimension that considers the average values of the latent variable scores.

The following two chapters deal with different concepts that enable researchers to model heterogeneous data. Chapter 4 first provides an overview of observed and unobserved heterogeneity, showing how disregarding heterogeneous data structures can provoke biased results. Next, we discuss measurement invariance, which is a primary concern before comparing groups of 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 allow accounting 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 5 introduces two 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.

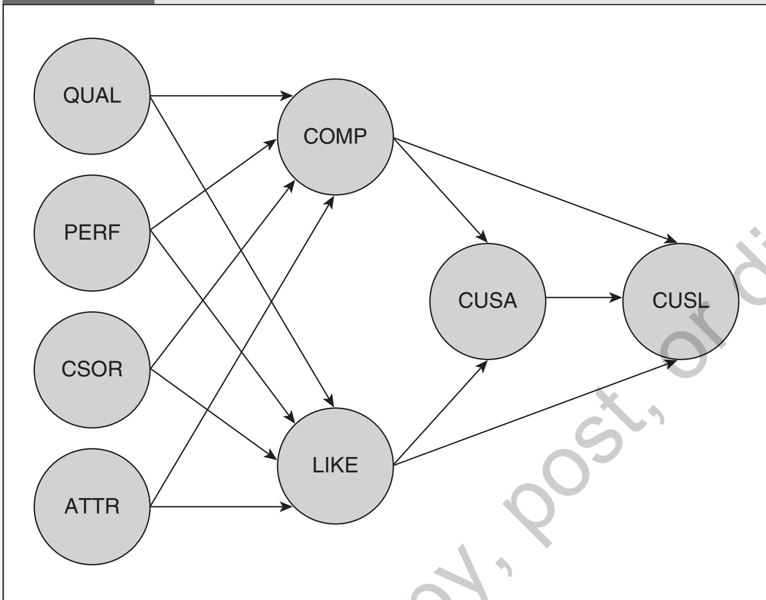
CASE STUDY ILLUSTRATION

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 you to do that. The example is drawn from a series of published studies on corporate reputation, which is general enough to be understood by many different areas of social science research, 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 et al. (2017) use in their primer on PLS-SEM. The model's purpose is to explain the effects of corporate reputation on customer satisfaction (*CUSA*) and, ultimately, customer loyalty (*CUSL*). Corporate reputation represents a company's overall evaluation by its stakeholders (Helm, Eggert, & Garnefeld, 2010). 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*). Schwaiger further identifies four antecedent dimensions of reputation: quality (*QUAL*), performance (*PERF*), attractiveness (*ATTR*), and corporate social responsibility (*CSOR*). Exhibit 1.6 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. Note that you should generally avoid using single items, particularly in PLS-SEM analyses (e.g., Diamantopoulos, Sarstedt, Fuchs, Kaiser, & Wilczynski, 2012; Sarstedt, Diamantopoulos, & Salzberger, 2016; Sarstedt, Diamantopoulos, Salzberger, & Baumgartner, 2016). However, we included this single item for illustrative purposes. Finally, the four exogenous constructs are measured by a total of 21 formative indicators. Exhibit 1.7 provides an overview of all items and item wordings. Respondents rated the questions on 7-point Likert

Exhibit 1.6 Corporate Reputation Model



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. Satisfaction and loyalty were measured with respect to the respondents' own service providers.

Exhibit 1.7 Items and Item Wordings

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 likeable company.

(Continued)

Exhibit 1.7 (Continued)

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] as my mobile phone services provider.
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.
qual_2	[The company] is an innovator, rather than an imitator with respect to [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, buildings, shops, etc.).

The measurement approach has been validated in different countries and applied in various research studies (e.g., Eberl & Schwaiger, 2005; Raithel & Schwaiger, 2014; Raithel, Wilczynski, Schloderer, & Schwaiger, 2010; Schloderer, Sarstedt, & Ringle, 2014). Research has shown that, compared to alternative reputation measures, the approach performs favorably in terms of convergent validity and predictive validity (Sarstedt et al., 2013). The data set used for all analyses in this book stems from Hair et al. (2017) and has 344 observations.

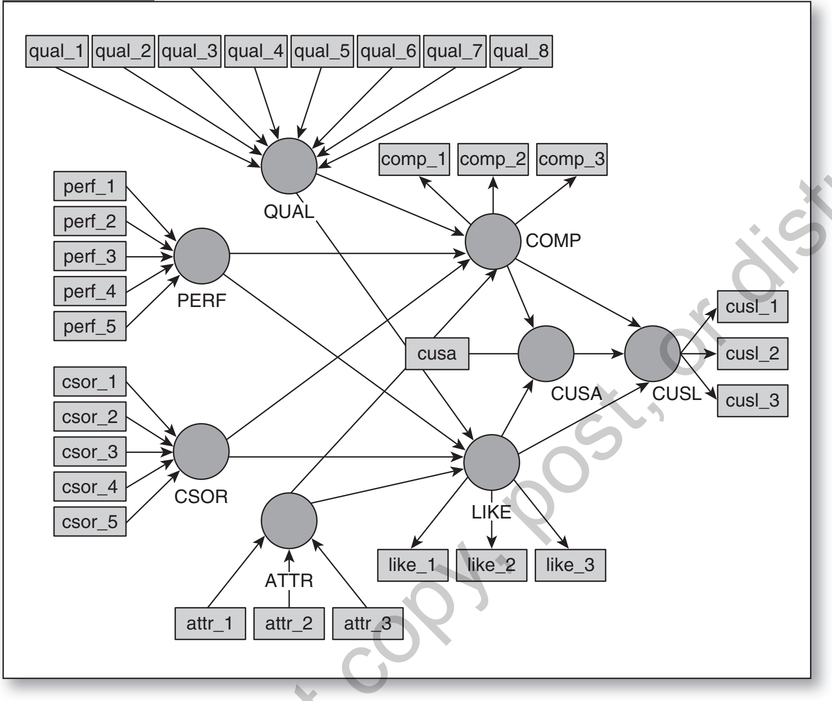
PLS-SEM Software

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 PLS software. Compared to LVPLS, which required the user to enter commands via a text editor, PLS-Graph represents a significant improvement, especially in terms of user-friendliness. However, PLS-Graph has not been further developed in recent years. The same holds for several other early software programs such as VisualPLS (Fu, 2006) and SPAD-PLS (Test&Go, 2006); see Temme, Kreis, and Hildebrandt (2010) for a review. 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 XLSTAT's PLSPM package, Adanco (Henseler & Dijkstra, 2016); PLS-GUI (Hubona, 2015); and particularly SmartPLS (Ringle et al., 2015; Ringle et al., 2005). Finally, users with experience in the statistical software environment R can also draw on packages such as semPLS (Monecke & Leisch, 2012) and plspm (Sánchez, Trinchera, & Russolillo, 2015), which facilitate flexible analysis of PLS path models. To date, SmartPLS 3 is the most comprehensive and advanced program in the field and serves as the basis for all case study examples in this book. The student version of the software is available free of charge at www.smartpls.com. The student version 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), you should use the professional version of SmartPLS, which is available as a 30-day trial version at www.smartpls.com. After the trial period, a license fee applies. Licenses are available for different periods of time (e.g., 2 months, 1 year, 3 years) and can be purchased through the SmartPLS website. The SmartPLS website includes a download area for the software, including the old SmartPLS 2 (Ringle et al., 2005) version, and many additional resources such as short explanations on 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 allows you to discuss PLS-SEM topics with other users.

Setting Up the Model in SmartPLS

Before you draw your model in SmartPLS 3, you need to have data that serve as the basis for running the model. The data we use with the reputation model can be downloaded either as comma-separated value (.csv) or text (.txt) data sets in the download section at www.pls-sem.com. Click on **Save Target As . . .** to save the data to a folder on your hard drive and then **Close.** SmartPLS can use both data file formats (i.e., .csv or .txt). Follow the onscreen instructions to save one

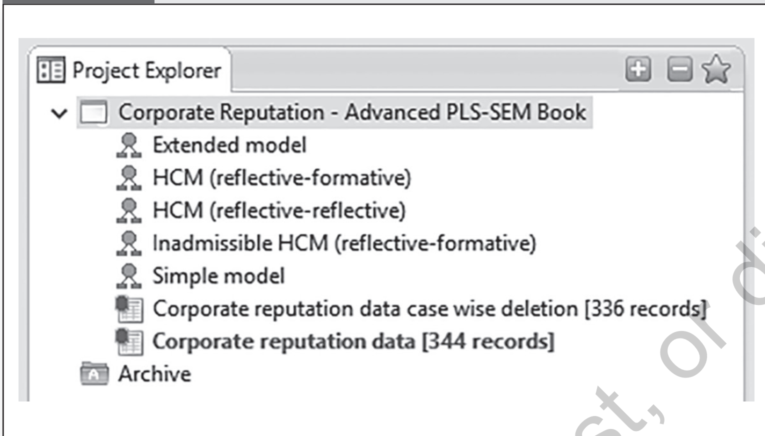
Exhibit 1.8 Corporate Reputation Model in SmartPLS



of these two files on your hard drive. With this data, as explained in Chapter 2 of the book by Hair et al. (2017), you can create the PLS path model as shown in Exhibit 1.8.

Alternatively, you can download the ready-to-use corporate reputation project file (**Corporate Reputation–Advanced PLS-SEM Book.zip**) for SmartPLS from www.pls-sem.com. Save the project file on your computer (or other device). Now run the SmartPLS 3 software by clicking on the desktop icon that is available after the software installation on your computer. Another possibility is to go to the folder where you installed the SmartPLS software on your computer. Click on the file that runs SmartPLS and then on the **Run** tab to start the software. To import the corporate reputation project file into the SmartPLS software, use the **File** → **Import Project from Backup File** option in the SmartPLS menu bar. Then you'll see the corporate reputation project in the SmartPLS **Project Explorer**. Unfold the project as shown in Exhibit 1.9 and double-click on the **Extended model**. Then the PLS path model as shown in Exhibit 1.8 appears in the SmartPLS modeling window.

Exhibit 1.9 The SmartPLS Project Explorer



Following the systematic procedure for applying PLS-SEM presented in Hair et al. (2017), the next steps entail the evaluation of the reflective and formative 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 et al. 2017) for a detailed discussion and illustration of these analysis steps. The case study illustrations in the following chapters will depart from here, assuming that the quality of the original model's measurement and structural models has been established.

SUMMARY

- **Understand the origins and evolution of PLS-SEM.** The precursors to the PLS-SEM method were two iterative procedures (i.e., principle component regression and PLS-R) that used least squares estimation to develop solutions for single and multicomponent models and for canonical correlation. 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 latent variables. In the 1980s Herman Wold proposed his "soft model basic design" underlying PLS-SEM as an alternative to CB-SEM, which has been labeled as hard modeling because of its much more rigorous assumptions in terms of data distribution and sample size (e.g., Falk & Miller, 1992). While both approaches were developed about the same time, CB-SEM became much more widely applied because of its early availability through the LISREL software since the late 1970s. It was not until the debut of Wynne Chin's PLS-Graph software in the mid-1990s that PLS-SEM began developing. With the release of SmartPLS 2 in 2005, PLS-SEM use grew exponentially.

- **Comprehend the principles of and recent developments in measurement.** Conceptual variables represent broad ideas or thoughts about abstract concepts that researchers establish and propose to measure in their research by means of constructs. Based on a construct definition, measurement models express how to measure the construct by means of a set of indicators. Measurement models can be specified reflectively, using effect 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 contends that composite indicators can be used to measure any type of property to which the focal concept refers, including attitudes, perceptions, and behavioral intentions. All measures of conceptual variables are approximations of or proxies for conceptual variables, independent from how they were derived. Thus, irrespective of the quality with which a conceptual variable is theoretically substantiated and operationally defined, and the rigor that encompasses measurement model development, because latent variables stem from data that are inherently imperfect, any measurement in structural equation models produces only proxies for latent variables.

- **Get to know the essential differences between PLS-SEM and CB-SEM and understand when to use each method.** PLS-SEM emphasizes prediction while simultaneously relaxing the demands regarding the data and specification of relationships. PLS-SEM maximizes the endogenous latent variables' explained variance by estimating partial model relationships in an iterative sequence of ordinary least squares regressions. In contrast, CB-SEM estimates model parameters so that the discrepancy between the estimated and sample covariance matrices is minimized.

A crucial conceptual difference between PLS-SEM and CB-SEM relates to the way each method treats the latent variables 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 represent the constructs. Importantly, the proxies produced by PLS-SEM are not assumed to be identical to the constructs that they measure—just as it is the case with construct measures produced by CB-SEM. PLS-SEM is not constrained by identification issues, even if the model becomes complex—a situation that typically restricts CB-SEM use—and does not require accounting for most distributional assumptions. Moreover, PLS-SEM can better handle formatively specified measurement models. Researchers should consider the two SEM approaches as complementary and apply the SEM technique that best suits their research objective, data characteristics, and model setup.

REVIEW QUESTIONS

1. Who developed the generalized PLS-SEM algorithm and what was the intention behind its development?
2. What is the difference between PLS-R and PLS-SEM?
3. What is the difference between common factor models and composite models?
4. What is the difference between reflective and formative measurement models?
5. What is PLSc and what was the intention behind its development?

CRITICAL THINKING QUESTIONS

1. Why is PLS-SEM the preferred method over CB-SEM for prediction?

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. What are the benefits of using PLSc over CB-SEM and PLS-SEM?

KEY TERMS

Attribute
Average variance extracted (AVE)
Causal indicators
CB-SEM
Common factor model approach
Common variance
Composite indicators
Composite model approach
Conceptual variables
Consistent PLS (PLSc)
Constructs
Construct definition
Covariance-based structural equation modeling (CB-SEM)
Effect indicators
Error variance
Factor-based SEM
Factor (score) indeterminacy
Focal object
Formative measurement model
Indicators
Items
Latent variable
Latent variable scores
Manifest variables
Measurement model
Mimic model
Multiple indicators and multiple causes (MIMIC) model
Partial least squares algorithm
Partial least squares path modeling
Partial least squares regression (PLS-R)
Partial least squares structural equation modeling
PLS algorithm
PLSc
PLS-R
PLS-SEM
Principal components regression
Reflective indicators
Reflective measurement model
Specific variance
Theoretical model
Unique variance

SUGGESTED READINGS

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

Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*. Advance online publication. doi:10.1037/met0000056

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Thousand Oaks, CA: Sage.

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning, 45*, 341–358.

Rigdon, E. E. (2016). Choosing PLS path modeling as analytical method in European management research: A realist perspective. *Long Range Planning*, *34*, 598–605.

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*, 3998–4010.