European management research using partial least squares structural equation modeling (PLS-SEM)

Keywords:
Management research
Partial least squares (PLS)
Structural equation modeling (SEM)
Path modeling
Quantitative methods
Method development
Empirical application

Hair, Sarstedt, Pieper, and Ringle's (2012) review study shows that partial least squares structural equation modeling (PLS-SEM) has become an increasingly applied multivariate analysis technique in management research. More recently, Richter, Sinkovics, Ringle, and Schlägel (2016) echo this result by showing that the number of PLS-SEM applications in (international) business research has increased substantially in the past few years. However, PLS-SEM is still new to many researchers who want to know: What exactly is PLS-SEM?

Most explanations limit themselves to the algorithm's statistical elucidations (e.g., Rigdon, 2013; Tenenhaus, Esposito Vinzi, Chatelin, & Lauro, 2005; Wold, 1982), while a few others include additional descriptions, such as PLS-SEM's historical background (e.g., Chin, 1998; Dijkstra, 2010, 2014; Lohmöller, 1989; Rigdon, 2012, 2014). Herman O. A. Wold (2006), the originator of the method, characterizes PLS-SEM as an "epoch-making 1960s innovation" that combines econometric prediction with the psychometric modeling of latent variables (also referred to as constructs), which multiple indicators (also referred to as manifest variables) determine.

To provide a better understanding of the approach, Fig. 1 shows a simple PLS path model with four latent variables, Y_1 to Y_4 (represented by circles), determined as the weighted sum of their assigned indicators x (represented by the rectangles). In other words, in the measurement model (also called the outer model), a block of directly observable indicators represents each latent variable that is not directly observable. In the structural model (also called the inner model), the latent variables have pre-defined and theoretically/conceptually established relationships.

The goal of the PLS-SEM approach is to generate latent variable scores that jointly minimize the residuals of the ordinary least squares (OLS) regressions in the model (i.e., maximize the explanation). The resulting latent variable scores are unique and determine the case values of each observation (i.e., the algorithm provides determinate latent variable scores). They also make it possible to predict the indicators (x_7-x_{12}) of the endogenous or dependent latent variables in the structural model (Y_3 and Y_4).

In short, PLS-SEM is a variance-based method that estimates composites representing latent variables in path models. Hair, Hult, Ringle, and Sarstedt (2017), for example, provide additional explications of PLS-SEM, including details on how to create and estimate PLS path models and how to evaluate the results (also see Chin, 1998, 2010; Falk & Miller, 1992; Haenlein & Kaplan, 2004; Hair, Ringle, & Sarstedt, 2011; Henseler, Hubona, & Ray, 2016; Roldán & Sánchez-Franco, 2012; Tenenhaus et al., 2005).

An alternative perspective on the PLS-SEM method sees the exogenous or independent latent variables' indicators (i.e., x_1—x_6 on the left side of Fig. 1) as the data input layer and the endogenous or dependent latent variables' indicators (i.e., x_7-x_{12} on the right side of Fig. 1) as the data output layer. The latent variables and their relationships represent the structural model that connects the input and the output layer. While the input and output data change (e.g., across time, industries, companies, products, customers, and countries), the structural model and its latent variables represent the stable, theoretically/conceptually established contextual link between the observed data on the input and output sides. Based on the structural model, the goal of the analysis is to predict the output layer data by means of the input layer data. To this end, the variance-based PLS-SEM approach uses OLS regressions and a structural model with latent variables, as well as the latter's relationships between the input and the output layers. "Thanks to the explicit case values of latent variables and structural residuals, the predictive relevance of a soft model can be explored by Stone-Geisser's cross-validation test" (Wold, 1982, p. 53). Against this background, it is possible to position PLS-SEM between covariance-based SEM (CB-SEM) and machine learning. While the former only focuses on the relationships between theory testing and confirmatory/explanatory modeling, the latter focuses primarily on prediction. If the two approaches are two ends of a continuum, PLS-SEM—with its prediction-oriented goal and its theoretically/conceptually established structural model of latent variables and their relationships—is, as shown in Fig. 2, positioned between the two ends.

"Why should there be a difference between explaining and predicting? The answer lies in the fact that measurable data are not accurate representations of their underlying constructs. The operationalization of theories and constructs into statistical models and measurable data creates a disparity between the ability to explain phenomena at the conceptual level and the ability to

http://dx.doi.org/10.1016/j.emj.2016.08.001