Chapter 64
Using Data Labels to Discover Moderating Effects in PLS-Based Structural Equation Modeling

Ned Kock
Texas A&M, USA

ABSTRACT

PLS-based structural equation modeling is extensively used in e-collaboration research, as well as in many other fields of research. Two main types of exploratory analyses are frequently employed in the context of PLS-based structural equation modeling: covariance (or correlation) analyses, where the covariances (or correlations) among all variables are inspected; and model-driven exploratory analyses, where one or more variations of theory-supported models are built and adjusted associations among variables are inspected. These analyses, while useful, provide limited insights about the possible presence of moderating effects. We discuss a general approach through which researchers can employ data labels, implemented as symbols that are displayed together with legends on graphs, to uncover moderating relationships among variables. The discussion is illustrated with the software WarpPLS version 4.0. While the approach is illustrated in the context of e-collaboration research, it arguably applies to any field where PLS-based structural equation modeling can be used.

INTRODUCTION

The method of path analysis (Wright, 1934; 1960) provides the foundation for structural equation modeling (SEM). Both path analysis and SEM rely on the development of models where causal assumptions are expressed through links among variables. In SEM variables that are measured indirectly through other variables are known as latent variables; the directly measured variables used to measure latent variables are known as indicators (Maruyama, 1998; Mueller, 1996). Often directly measured variables are obtained from questionnaires where answers are provided on numeric scales, such as Likert-type scales (Cohen et al., 2003).

DOI: 10.4018/978-1-5225-0196-1.ch064

Two main approaches to SEM exist today (Kock & Lynn, 2012; Schumacker & Lomax, 2004). In both of them researchers define a measurement model by linking indicators to their respective latent variables, and define a structural model by linking latent variables among themselves; both based on causal assumptions supported by theory. The classic approach is known as covariance-based SEM (Schumacker & Lomax, 2004). This approach estimates model coefficients (e.g., path coefficients) based on the minimization of differences between covariance matrices, and makes parametric assumptions in the calculation of coefficients that serve as a basis for calculation of significance levels (often expressed as P values). The more modern SEM approach is known as variance-based SEM or PLS-based SEM (Kock & Lynn, 2012), where "PLS" stands for "partial least squares". This approach estimates model coefficients by creating latent variable scores through weighted aggregations of indicators. It makes no parametric assumptions in the calculation of P values; for example, it does not assume that latent variable or indicator scores are normally distributed.

PLS-based SEM is extensively used in e-collaboration research and in related fields, such as management information systems (Guo et al., 2011; Kock, 2013a; Kock & Lynn, 2012; Schmiedel et al., 2014). These are fields where the topic of statistical moderation is given a central role in empirical research (Carte & Russel, 2003). In a linear model, a variable X is said to moderate a linear relationship between two other variables Y and Z if a variation in X leads to a variation in the slope of the best-fitting line for the relationship between Y and Z. Frequently X represents elements of the physical and social contexts in which Y affects Z. Generally speaking, those contexts play key roles in how e-collaboration phenomena unfold (Kock, 2005).

In this article we discuss how researchers can employ data labels, implemented as symbolic elements that are displayed together with legends on a graph, to identify moderating relationships among variables. Our discussion focuses primarily on linear models, but can be extended to models that include nonlinear relationships. The software WarpPLS version 4.0 (Kock, 2013b) is used in this discussion. The extensive set of results produced by this software allows us to illustrate the several steps involved in using data labels to uncover the existence of statistical moderation in an exploratory analysis and subsequently model the moderation explicitly in a confirmatory analysis.

DATA USED IN THE ANALYSIS

We created 300 rows of data, equivalent to 300 returned questionnaires, for several latent variables and indicators based on a Monte Carlo simulation (Robert & Casella, 2005; Paxton et al., 2001). Using this method we departed from a "true" model, which is a model for which we know the nature and magnitude of all of the relationships among variables beforehand. The true model was based on an actual study of the effects of e-collaboration technology use on team-based project success, which was previously used by Kock & Lynn (2012) to illustrate their discussion of vertical and lateral collinearity. At the time of this writing this data set was publicly available from the WarpPLS web site: http://warppls.com

The following variables were included in this analysis: e-collaboration technology use (ECollab), project management (Projmgt), job satisfaction (JSat), and project success (Success). E-collaboration technology use (ECollab), the main technology-related variable in the model, measures the extent to which a team that is tasked with the development of a new product (e.g., a new toothpaste or airplane part) uses an e-collaboration technology that integrates several synchronous and asynchronous features (e.g., e-mail, text-based chat, video-conferencing, discussion board).