Journal of Family Business Strategy 5 (2014) 116–128

Contents lists available at ScienceDirect

Journal of Family Business Strategy

journal homepage: www.elsevier.com/locate/jfbs

A comparative study of CB-SEM and PLS-SEM for theory development in family firm research

Claudia Binz Astrachan [a], Vijay K. Patel [b], Gabrielle Wanzenried [c,*]

[a] Witten/Herdecke University, Germany
[b] Kennesaw State University, USA
[c] Lucerne University of Applied Sciences and Arts, Switzerland

ARTICLE INFO

Keywords:
Structural equation modeling (SEM)
Covariance-based SEM
Partial least squares-SEM
Family firms
Organizational reputation
Organizational trustworthiness

ABSTRACT

Structural equation modeling (SEM) has become the methodology of choice for many family business researchers investigating complex relationships between latent constructs, such as family harmony or family cohesion. Its capability to evaluate complex measurement models and structural paths involving a multitude of variables and levels of constructs has enabled family business researchers to investigate complex and intricate relationships that previously could not be easily untangled and examined. In many cases, however, researchers struggle to meet some of the challenging requirements of covariance-based SEM (CB-SEM), the most commonly used approach to SEM, such as distribution assumptions or sample size. In this article, we point out the benefits and disadvantages of CB-SEM, and present a comparison with partial least squares-SEM (PLS-SEM) using an identical sample. We find that even though both methods analyze measurement theory and structural path models, there are many advantages in applying PLS-SEM.

1. Structural equation modeling in a nutshell

Structural equation modeling (SEM) has seen a dramatic rise in attention and utilization across a variety of scientific disciplines such as strategic management (Shook, Ketchen, Cycyota, & Crockett, 2003), marketing (Chin, Peterson, & Brown, 2008) and psychology (MacCallum & Austin, 2000) over the last decade (Hair, Ringle, & Sarstedt, 2011b). Statistically, SEM represents an advanced version of general linear modeling procedures (e.g., multiple regression analysis), and is used to assess "*whether a hypothesized model is consistent with the data collected to reflect [the] theory*" (Lei & Wu, 2007, p. 34). While SEM is a general term encompassing a variety of statistical models, covariance-based SEM (CB-SEM) is the more widely used approach in SEM, and many researchers simply refer to CB-SEM as SEM. This reference is naïve, however, because partial least squares (PLS) is also a useful and increasingly applied approach to examine structural equation models (Hair, Sarstedt, Ringle, & Mena, 2012).

Structural equation modeling is a multivariate analytical approach used to simultaneously test and estimate complex causal relationships among variables, even when the relationships are hypothetical, or not directly observable (Williams, Vandenberg, & Edwards, 2009). Concurrently combining factor analysis and linear regression models, SEM allows the researcher to statistically examine the relationships between theory-based latent variables and their indicator variables by measuring directly observable indicator variables (Hair, Hult, Ringle, & Sarstedt, 2014). While SEM is similar to multiple regression in the sense that both techniques test relationships between variables, SEM is able to simultaneously examine multi-level dependence relationships, "*where a dependent variable becomes an independent variable in subsequent relationships within the same analysis*" (Shook, Ketchen, Hult, & Kacmar, 2004, p. 397) as well as relationships between multiple dependent variables (Jöreskog, Sörbom, du Toit, & du Toit, 1999).

The objective of this article is to evaluate the benefits and limitations of SEM in general, and in family business research in particular, by directly comparing two major approaches to structural modeling – covariance based SEM (CB-SEM) and variance-based SEM (PLS-SEM) (Sarstedt, Ringle, Smith, Reams, & Hair, 2014; Sharma & Kim, 2013). While CB-SEM and PLS-SEM are two different approaches to the same problem – namely, the analysis of "*cause–effect relations between latent constructs*" (Hair, Ringle, & Sarstedt, 2011a, p. 139), they differ not only in terms of their basic assumptions and outcomes, but also in terms of their estimation procedures (Hair et al., 2014; Shook et al., 2004). PLS-SEM uses a regression-based ordinary least squares (OLS) estimation method with the goal of explaining the latent constructs' variance by "*minimizing the error terms [and maximizing] the R^2 values of the*

* Corresponding author. Tel.: +1 404 242 0803.
E-mail address: claudia.astrachan@gmail.com (G. Wanzenried).

1877-8585/$ – see front matter
http://dx.doi.org/10.1016/j.jfbs.2013.12.002

(target) endogenous constructs" (Hair et al., 2014, p. 14; Ringle, Sarstedt, Hair, & Pieper, 2012). CB-SEM, on the other hand, follows a maximum likelihood (ML) estimation procedure and aims at "*reproducing the covariance matrix [i.e., minimizing the difference between the observed and estimated covariance matrix], without focusing on explained variance*" (Hair et al., 2011a, p. 139). In other words, with CB-SEM, the R^2 is a by-product of the overall statistical objective of achieving good model fit (Hair et al., 2014).

Using a sample of 253 Swiss consumers surveyed in 2012 evaluating the effects of corporate expectations on the perceived level of expertise and trustworthiness of family-owned companies, we apply both CB-SEM and PLS-SEM to analyze the data. This approach enables us to not only compare the requirements of each method, the way in which the models are specified, and the applicability and user-friendliness of available software, but also the results and interpretations.

The remainder of this article is structured as follows: first, we briefly highlight the most important benefits of SEM. We then summarize the results of several important articles in family business research that utilized SEM, and point out how SEM contributed to the findings of these studies. Third, the research context of the example used in this study is briefly described, and the hypotheses as well as an outline of the methodology are presented. Fourth, we discuss the results from the CB-SEM and PLS-SEM analyses. Finally, practical observations and conclusions are provided, and limitations and suggestions for further research are presented.

2. The benefits and limitations of SEM

2.1. The benefits of SEM

The question of why researchers might want to use SEM is quite simple. The process of applying SEM enables researchers to more effectively evaluate measurement models and structural paths, particularly when the structural model involves multiple dependent variables, latent constructs based on multi-item indicator variables, and multiple stages/levels of constructs in a structural model. While there are many reasons to use SEM in social sciences research, we consider the following to be the most relevant.

When dealing with latent constructs and complex models: Many constructs investigated in the social sciences are latent constructs that cannot be observed, or measured directly. Examples include family influence and family cohesion. Moreover, especially at the theory development and testing stages there may be multiple constructs and interactive effects resulting in a complex model. While a latent construct may be measurable to some extent by means of a directly observable indicator variable (e.g., degree of family ownership, number of family members in management), these indicator measures may not reflect the latent variable entirely accurately, which means the measurement will contain error as will the results. By explicitly assessing error in the structural model, SEM "*provides a powerful means of simultaneously assessing the quality of measurement and examining causal relationships among constructs*" (Wang & Wang, 2012, p. 1). So while multiple regression analysis assumes there is no error in the data, SEM recognizes and accounts for the error in each measured item in an effort to improve the accuracy of findings. Additionally, the SEM approach is designed to consider interactive effects and complex models to find an optimal model that reduces cross-loadings and identifies the higher loadings for relevant measures.

When analyzing direct, indirect, and total effects: SEM facilitates the assessment of direct, indirect and total effects. Direct effects include relationships between independent and dependent variables, e.g., family ownership has a direct positive effect on firm performance. Indirect effects involve relationships between independent and dependent variables that are mediated or moderated by some other variable, e.g., the effect of family ownership on firm performance is moderated by the owning family's involvement in management. Total effects relate to the sum of two or more direct or indirect effects. In comparison to other statistical procedures such as regression, SEM enables researchers to not only simultaneously assess the relationships between multi-item constructs, but also to reduce the overall error associated with the model. In contrast to multiple regression analysis, which cannot directly deal with the measurement issues of multi-item constructs, SEM is specifically designed to improve multi-item measurement models by directly accounting for error.

When assessing structural models: While regression also allows researchers to evaluate structural relationships using path analysis (examining each path separately), SEM facilitates simultaneous analysis of all structural relationships (i.e., relationships or paths among numerous variables, e.g., family ownership, family cohesion and performance), and is an inherently simpler approach that leads to more accurate results. CB-SEM and PLS-SEM use different approaches when assessing the quality of a structural model. For example, with CB-SEM fit is based on accurately estimating the observed covariance matrix, while with PLS-SEM fit is based upon accounting for explained variance in the endogenous constructs (Hair et al., 2014). As a result of model fit requirements, however, CB-SEM often eliminates relevant indicator variables, thereby reducing the validity of constructs. In contrast, PLS-SEM creates composite constructs that generally include additional theory-based indicator variables (Rigdon, 2012), while still optimizing predictive accuracy and relevance. Also, PLS-SEM analyses can easily incorporate single-item measures, and can obtain solutions to much more highly complex models, i.e., models with a large number of constructs, indicators and structural relationships (Hair et al., 2014; Ringle, Sarstedt, & Hair, 2013).

2.2. The limitations of SEM

The fact that modern SEM software (such as AMOS, LISREL and SmartPLS) does not require profound statistical knowledge has made investigation of complex statistical problems accessible to non-statisticians (Babin, Hair, & Boles, 2008; Hair, Black, Babin, & Anderson, 2010). Yet, while ease of access to SEM has increased the number of meaningful and valuable contributions, recent reviews of SEM applications provide grounds for criticism of methodological flaws and shortcomings in the execution of SEM in many contributions (e.g., Hair et al., 2012; Williams et al., 2009). Being a highly sophisticated statistical tool, "*insight and judgment are crucial elements of its use*" (Shook et al., 2004, p. 397). Thus, to obtain meaningful and valid results it is essential to understand when it is appropriate to use SEM, its requirements and interpretation, and also the potential trade-offs when compared to other methods.

When unable to correctly identify a research model: In the case of CB-SEM in particular, since it is a confirmatory approach, the method requires the specification of the full theoretical model prior to data analysis. The researcher(s) must therefore define the exact number of dependent (endogenous) and independent (exogenous) variables used in the theoretical model, the relationships between these latent variables, the type of measurement model (formative or reflective), and the number of indicator variables required to ensure a valid and reliable measure of all constructs (e.g., Williams et al., 2009). Only when a model is correctly specified can all parameters be estimated (Lei & Wu, 2007). Thus, if the model lacks a sound theoretical foundation, and if the direction of the relationship between variables cannot be determined, CB-SEM should not be the method of choice. In contrast, PLS-SEM, which is particularly suitable for early-stage