A Review of PLS-SEM as Statistical Approach for Business Research

Dr Sarvesh Kumar
Assistant Professor,
School of Commerce and Management Studies,
Central University of Himachal Pradesh,
Dhauladhar campus— II, Dharamshala,
Himachal Pradesh

Vikas Kumar Tyagi
Assistant Professor,
Department of Management Studies,
Panipat Institute of Engineering and Technology,
Samalkha, Haryana

Dr. Yoginder Singh Kataria
Professor,
Faculty of Commerce and Management
SGT University, Gurugram, Haryana

Abstract

Structural Equation Modeling (SEM), a Second generation multivariate data analysis technique is in use for data analysis, especially hypothesis testing from a long time. There are different approaches to SEM used in research. The objective of this research paper was to review different Structural Equation Modelling techniques used in Business Research, with special reference to variance based Partial Least Square Structural Equation Modeling (PLS-SEM). PLS-SEM is more sensitive, simpler, and powerful statistical technique for data analysis. It is mainly used in theory development for exploratory purposes and has less strict assumptions. Several types of research have been done in the past on this subject; this paper compiles all the ideas by reviewing them and covering them in the least complicated form. It would especially help the researchers and practitioners in understanding the differences between different approaches to SEM and its applications.

KEYWORDS: Structural Equation Modeling, Partial Least Square, Covariance, Variance.

JEL CLASSIFICATION: C14, C31, C88

Introduction

Volatility, Uncertainty, Complexity, and Ambiguity (VUCA) originated in the US Military (Whiteman, 1998) as cited by Bennett & Lemoine (2014). It has become a synonym of constant change in dynamic business research scenario. Embedding mathematical analysis simplifies decision making in a dynamic business environment. PLS-SEM has evolved as a handy tool for researchers. This tool also helps researchers and decision makers to reach confident decisions concerning their defined problems. Sewall Wright first developed Path analysis models in the year 1921 (Wolfe, 1980). Structural Equation Modeling (SEM) is a second generation multivariate data analysis technique (Elangovan & Rajendran, 2015). SEM has advantage of analyzing multiple layers of links between independent (IV) and dependent variables (DV) simultaneously over first generation regression models like Linear Regression, Analysis of variance (ANOVA), and Multivariate analysis of variance (MANOVA); it can. Researchers use SEM extensively for hypothesis testing (Bagozzi & Yi, 1988).
There are several, but the researchers most commonly use two approaches to SEM. The first approach is the more widely used and is older than PLS-SEM; it is Covariance-based SEM (CB-SEM), which is more of confirmatory and conclusive. Covariance is the extent of how much two variables change together or how well the variables are jointly related (Davis & Pecar, 2013). The second approach, Variance based Partial Least Squares SEM (PLS-SEM) is more exploratory in nature and was less commonly used till recently. Variance is the measure of the dispersion of the observations, to check how dispersed the data values are about the mean values (Davis & Pecar, 2013). Hwang & Takane, (2004) came up with the third approach, i.e. Generalized Structured Component Analysis (GSCA), after these two methods. The researcher has kept GSCA outside the purview of this research.

In the quest of statistical significance, during different time horizons, different types of methods to interpret data have evolved, solving the problem of that time. Over the years, the inherent gaps of existing methodology paved the path of newer methodology whose structure has been built by refining current methods. Considering one technique to be better than the other would not be right, both have their advantages and disadvantages, discussed later in this paper. There are several studies done on this subject, giving their opinions and different aspects of the subject; this paper compiles all the ideas by reviewing them. This research paper would help in understanding the differences between the concepts. This paper is an essential guide reducing complexity, for learning it in more details researcher suggests to go through research such as Gaskin (2018); Hair Jr, Hult, Ringle, & Sarstedt (2016) and Byrne, (2016).

In VUCA times PLS-SEM is gaining lot of prominence in business research as popular choice of business research methods. Internationally much work has been done by using the concept PLS-SEM, such as studies done in different fields such as by Aryanto, Fontana, & Afiff (2015); Astrachan, Patel, & Wanzenried (2014); Hair, Sarstedt, Pieper, & Ringle (2012); Okazaki, Mueller, & Taylor (2010); and Schweiger, Sarstedt, & Taylor (2010) in the area of Operations Management. Chen, Preston, & Xia (2013) and Hoffmann, Schiele, & Krabbendam (2013) in the field of Supply Chain Management. Kallunki, Laitinen, & Silvola (2011) in Accounting Research. In India- Atulkar & Kesari (2018); Kesari, B., & Atulkar, S. (2016); Kamath, Rodrigues, & Desai (2016); Shanmugapriya, & Subramanian (2015); Venkatesh, Sykes, & Venkatraman (2014); Seetharaman, Bajaj, Raj, & Saravanan (2013); did studies by using PLS-SEM in the area of business research.

In this paper, the introduction is followed by the comparison between both PLS-SEM and CB-SEM and followed by the advantages of one over another and situations when each one of them should be selected. In the end, the conclusory remarks about the issue are mentioned.

Comparative Analysis between PLS-SEM and CB-SEM

Karl Joreskog developed CB-SEM (Joreskog, 1970) in behavioural sciences, whereas PLS-SEM was developed initially by Herman Wold (1974) in social sciences. Kosik, Kahai, & Piovoso, (2009) considered CB-SEM as hard modelling for theory testing whereas considered PLS-SEM as soft modelling approach for the theory development. Tenenhaus (2008) called the PLS-SEM as component based SEM, just as the other one is called Covariance-based approach.

The most significant difference between them lies in their respective objectives for which they are used. CB-SEM helps in involving theoretical covariance matrix by estimating model parameters with an objective of minimizing the differences between theoretical covariance matrix and the estimated covariance matrix, without focusing on the explained variance. Whereas, PLS-SEM is used with an objective of maximizing the explained variance of the dependent latent construct in business research (Hair, Ringle, & Sarstedt, 2011). This is the reason why majority of researchers do not accept standard goodness-of-fit statistic of PLS-SEM.

CB-SEM is the more popular technique as it provides flexibility of using various software's like Mplus, EQS, LISREL, and AMOS. Whereas, PLS-SEM is a less popular technique despite being a more robust estimator (Reinartz, Haenlein, & Henseler, 2009). The main reason for its lower popularity is also because the software on which it runs like: SmartPLS by Ringle, Wende, and Will in 2005 and PLS-Graph by Chin in 2003 were developed much later. Other PLS Software’s such as VisualPLS, WarpPLS, and ‘R’ statistical software package, can also be used to run PLS-SEM. Although VisualPLS and PLS-Graph have graphical interfaces they have not received any significant updates since their release, and ‘R’ requires a bit of advanced programming language skills, making SmartPLS most used and popular software for PLS applications. Its 2.0 version is available freely, making it the best-suited software for the purpose (Wong, 2013).
### Table 1: Studies comparing CB-SEM vs PLS-SEM, performed in different fields:

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Area of Research</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richter, Sinkovics, Ringle, &amp; Schlaegel (2016)</td>
<td>International Business Research</td>
<td>Out of 424 studies reviewed which used SEM, 379 were using CB –SEM, and the rest 45 were using PLS-SEM. They were used due to their lower sample sizes and data measurement issues in place of their objectives. Studies still don’t reap the benefits of PLS -SEM to its full Extent.</td>
</tr>
<tr>
<td>Kaufmann &amp; Gaeckler (2015)</td>
<td>Supply Chain Management</td>
<td>Use of PLS -SEM has magnified in the SCM recently, but most of them didn’t follow the standards of the technique. They found CB -SEM to be better for the subject if its assumptions are met.</td>
</tr>
<tr>
<td>Astrachan, Patel, &amp; Wanzenried (2014)</td>
<td>Family Firms Research</td>
<td>Found PLS -SEM better than CB -SEM for the studies in their area of research. PLS -SEM enables the extension of more indicator variables, whereas, CB-SEM explains variance better, but in the case of Non-normal data, CB-SEM gives inflated $R^2$.</td>
</tr>
<tr>
<td>Hair, Sarstedt, Ringle, &amp; Mena (2012)</td>
<td>Marketing Research</td>
<td>PLS-SEM has become more widely used in Marketing Research. However, this has been misunderstood as it lacks in the standard textbooks. It is a robust technique but should not be applied to ditch the assumptions of CB-SEM.</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Field of Research</td>
<td>Findings</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Peng &amp; Lai (2012)</td>
<td>Operation Management Research</td>
<td>PLS-SEM is most widely used in the field of the information system, and not as widely in the area of Operation Management field, where still CB-SEM is used. They found CB-SEM to be superior, but only if its assumptions are met otherwise, PLS-SEM should be opted.</td>
</tr>
<tr>
<td>Gefen, Rigdon, &amp; Straub (2011)</td>
<td>Administrative and Social Science Research</td>
<td>Found that it would not be right to mention one of them to be better than another. However, they should be opted as per their objectives. PLS-SEM should be used in exploratory studies, whereas CB-SEM in confirmatory studies.</td>
</tr>
<tr>
<td>Lee, Petter, Fayard, &amp; Robinson (2011)</td>
<td>Accounting Research</td>
<td>PLS-SEM is commonly used in Social Sciences but not in Accounting discipline where traditional regression is used. Studies in this field are yet to avail its benefits to its full extent.</td>
</tr>
<tr>
<td>Fornell &amp; Bookstein (1982)</td>
<td>Marketing Applications</td>
<td>Found PLS-SEM to be a more feasible option. Although CB-SEM was found to give statistically precise results but need to follow strict assumptions and require a larger sample size for accurate results. Whereas, PLS-SEM have prediction accuracy even with smaller sample sizes and can also deal with larger models with many variables.</td>
</tr>
</tbody>
</table>

Source: Literature Review
Advantages of one over another

According to Hair, Ringle, & Sarsted (2011) PLS-SEM has a higher level of statistical power as compared to CB-SEM; generates better path coefficients and significance level and is more sensitive in detecting relationships (Sosik, Kahai, & Piovoso, 2009); is simpler in nature (Tenenhaus, 2008), yet can deal with high model complexity (Hair, Ringle, & Sarstedt, 2011). PLS techniques can work even when there are just one or two items per construct, unlike CB-SEM. It can deal with both formative and reflective measurements models more easily than CB-SEM can, which requires relatively more complex rules (Hair, Ringle, & Sarstedt, 2011 and Sosik, Kahai, & Piovoso, 2009).

PLS methods are non-parametric techniques (Nagarajan, Savitskie, Ranganathan, Sen, & Alexandrov, 2013), unlike Co-variance based methods, which are parametric. Therefore, unlike CB Techniques, researchers do not have to satisfy any sets of assumptions before the application of PLS techniques. Model specifications or data in PLS-SEM do not use any limiting assumptions. While, multivariate normality of data can work well on non-normal data; PLS algorithm adjusts a non-normal data according to the central limit theory (Cassel, Hackl, & Westlund, 1999); minimum sample size as it can work with a small and much wider range of sample sizes (Diamantopoulos & Siguaw, 2013 and Ringle, Sarstedt, & Straub, 2012). Observational independence and interval scaled data as PLS-SEM can be applied even when the data is non-independent or is in Ordinal or Nominal scale (Sosik, Kahai, & Piovoso, 2009). According to Wong (2013), it is challenging to find the data that meet all these assumptions. Moreover, as commented by Wold (1982) as cited in Hair, Ringle, & Sarstedt (2011), the informational and distributional requirements or assumptions for CB-SEM are unrealistic. They are making PLS Techniques more realistic.

PLS-SEM is more suitable on the smaller sample sizes compared to CB-SEM (Wong, 2013); (Sosik, Kahai, & Piovoso, 2009); (Tenenhaus, 2008) and (Marcoulides & Saunders, 2006). Ringle, Sarstedt, & Straub (2012) reviewed/meta-analyzed 204 of studies which were done using PLS-SEM, they found that the average sample size in those studies was 238.12, and the median was 198. In addition to that, it was evident through the study that 33.8 studies used this method only because of having low sample sizes. Tenenhaus, Pages, Ambroisine, & Guinot (2005) even did research based on just six subjects.

Even though PLS-SEM has the edge over CB-SEM, CB-SEM also has few comparative advantages over PLS-SEM. PLS-SEM allows for only recursive relationships in the structural models with no causal loops allowed. Whereas, CB-SEM does not have any such restriction and can also work well with the non-recursive models; CB-SEM can deal with bidirectional relationships unlike in PLS-SEM where relationships tend to unidirectional only (Hair, Ringle, & Sarstedt, 2011). CB-SEM gives better global goodness of fit criterion (Hair, Ringle, & Sarstedt, 2011) compared to PLS-SEM with no global measure of goodness of model fit.

As PLS-SEM output does not give any overall model fit in, therefore Goodness of fit (GoF) proposed by Tenenhaus, Vinzi, Chatelin, & Lauro (2005) can be used to assess the structural model. The geometric mean (G.M.) of the Average Variance Extracted (AVE) and the average Coefficient of Determination (R²) is used for the PLS methods are non-parametric techniques (Nagarajan, Savitskie, Ranganathan, Sen, & Alexandrov, 2013); PLS Techniques more realistic. Ringle, Sarstedt, & Straub (2012) reviewed/meta-analyzed 204 of studies which were done using PLS-SEM, they found that the average sample size in those studies was 238.12, and the median was 198. In addition to that, it was evident through the study that 33.8 studies used this method only because of having low sample sizes. Tenenhaus, Pages, Ambroisine, & Guinot (2005) even did research based on just six subjects.

Selection between PLS-SEM Vs CB-SEM

As the researcher has already discussed the positive and negative aspects of both the methods in this study; now will discuss the situations in which they should be applied.

CB-SEM should be selected when the prior theory is strong and additional testing & validation are the research objectives. CB-SEM should also be applied when either the model is non-recursive, or when the assumptions of parametric tests are fulfilled, or when bidirectional paths are used in the model (Wong, 2013; and Hair, Ringle, & Sarstedt, 2011).

Whereas, PLS-SEM should be selected, if the research is more of exploratory in nature or if it is an extension to an existing theory. PLS-SEM should be used with the goal of predicting key target and driver constructs, i.e. more for theory development rather than theory confirmation (Hair, Ringle, & Sarstedt, 2011), notably when the sound theory base is missing. PLS-SEM should be used when available theories on the subject are insufficient (Wong, 2013). Especially during the early stage of theory development (Sosik, Kahai, & Piovoso, 2009). PLS-SEM is strongly prescribed when the model under study are complex or when formative constructs are part of the structural model (Ringle, Sarstedt, & Straub, 2012). Moreover, of course, its usage becomes must when the underlying assumptions of CB-SEM are violated (Reinartz, Haenlein, & Henseler, 2009), or when there is single item per construct (Ringle, Sarstedt, & Straub, 2012).
However, in the condition when the sample size is considerably large, then both the techniques give similar results, irrespective of the assumptions or anything else (Sosik, Kahai, & Piovoso, 2009). As according to the Figure: 1, PLS-SEM and Figure: 2, CB-SEM same manifest variables (USEF3, EOU3, BI3, ATT2, USE1) had maximum loadings for the Latent Variables (USEF, EOU, BI, ATT, USE). Also, Path EOUUSEF followed by USEFBI had maximum and second highest coefficient values, and the sequence of other paths were also the same in both the models. The results were the same; only the values were different. The reason for this could be because the dataset was huge as in this hypothetical it was of 1,190 responses.

Figure: 1 Partial least squares (PLS) results

Source: (SmartPLS, 2018)

Figure: 2 CB -SEM Maximum likelihood (ML) based results (AMOS; standardized coefficients)

Source: (SmartPLS, 2018)
Steps to be followed in PLS-SEM concerning reflective scale:

First of all, the objectives should be clearly defined and if they suit for PLS-SEM, then only it should be applied. Additionally, the assumption of the CB-SEM should be checked before proceeding further with the PLS-SEM if the data set fails to satisfy the assumptions than one would be forced to move towards the PLS Techniques. One should mention the reasons for using PLS Technique in their research.

There are two sub-models in a PLS Structural Equation Model; (Sosik, Kahai, & Piovoso, 2009). The first component of PLS-SEM model is the structural model or inner loop, having endogenous constructs which are explained by relationships in structural model with the other constructs (Hair, Ringle, & Sarstedt, 2011). It stimulates the relationships among the independent and dependent latent variables. The second component is an outer loop or measurement model having exogenous constructs with no structural path relationship between them (Hair, Ringle, & Sarstedt, 2011). It represents the relationships between the latent variables and their manifest variables.

It cannot be said that there are no restrictions on the sample size of the data set while performing PLS Techniques. Sample size in case of PLS-SEM should be at least equal to or more than ten times highest formative variables measuring one construct. (Hair, Ringle, & Sarstedt, 2011). Alternatively, minimum ten times the highest statistical paths directed at a latent construct in the structural model. According to Marcoulides & Saunders (2006) as cited in Wong (2013); minimum sample size should be 91 if maximum number of arrows pointing towards a latent variable is 10.

Checking validity and reliability is the most crucial step while performing PLS Techniques:

- For Internal consistency/reliability (Nunnally & Bernstein, 1994): in place of Cronbach alpha, Composite Reliability is the better method in PLS-SEM, as it does not assume all the indicators to be equally reliable (Bagozzi & Yi, 1988). Composite reliability of more than 0.60 is regarded as satisfactory.

- For indicator reliability (Hair, Ringle, & Sarstedt, 2011 and Bagozzi & Yi, 1988): Indicator loadings should be more than 0.70 and loadings that are less than 0.4 should not be included in reflective scale.

- For Convergent validity (Hair, Ringle, & Sarstedt, 2011) and (Bagozzi & Yi, 1988): Average Variance Extracted (AVE) should be more than 0.5, meaning that the latent variable explains more than half of the variance of its indicators.

- For Discriminant validity (Hair, Ringle, & Sarstedt, 2011): all cross-loadings should be lower than indicator loadings and AVE should be more than the construct's highest squared correlations with any other latent construct (Fornell-Larcker Criterion) (Fornell & Larcker, 1981).

Measuring 'Coefficient of determination' (R^2), i.e. model's predictive accuracy is primary evaluation criteria for the structural model. R^2 values of 0.2 are considered high in consumer behaviour studies. Values of 0.7, for endogenous latent construct is called as substantial, 0.5 as moderate and 0.25 as weak (Hair, Ringle, & Sarstedt, 2011). Additionally, PLS Algorithm must converge in maximum of 300 iterations, if it does not converge in 300 iterations it would have meant that data were abnormal due to the reasons such as the sample size could be too small, evidence of the existence of outliers, data having too many identical values in indicator and this would require further investigation (Wong, 2013).

Bootstrapping, which is a non-parametric method, allows for testing null hypothesis that a coefficient equals to zero, through this, the significance of the coefficient can be analyzed. Bootstrapping should be done with samples of at least 5,000 where number of cases should not be less than the number of original observations. Omission distance (d) should be chosen between 5 and 10. Critical t-values for a two-tail test are 1.65 (significance at the level of 10 percent), 1.96 (significance at the level of 5 percent), 2.58 (significance at the level of 1 percent) (Hair, Ringle, & Sarstedt, 2011).

Model's capacity to predict is significant (Rigdon, 2014) and can be measured through Stone-Geisser's (Q^2) (Stone, 1974) obtained through Blindfolding, which gives 'cross-validated redundancy' and 'cross-validated communalinity'. Hair, Ringle, & Sarstedt (2011) suggested using cross-validated redundancy (Q^2) of endogenous latent variable and its value to be more than zero for explaining the construct's predictive relevance.

With the significance, it is also f^2 value indicates the effect of the construct removed for a particular endogenous construct. The values of 0.02 represent small, 0.15 medium and 0.35 large effects (Cohen, 1988). If an exogenous construct strongly contributes to explaining an endogenous construct, the difference between R^2 included and R^2 excluded should be high, leading to high effect size (f^2 value). Essential to measure the magnitude of the influence, which can be done through Cohen's f^2 or Effect size.
\[ r^2 = \frac{(R^2_{AB} - R^2_A)}{1 - R^2_{AB}} \]

Where,

\[ R^2_A = \text{variance accounted for in the population by variable set A} \]

\[ R^2_{AB} = \text{Variance accounted for in the population by variable set A and B Together} \]

**Conclusion:**

Data itself is not the solution to any problem; it is the analysis of data which shows the decision path to any manager. For the analysis, several techniques could be used. Comparisons among the techniques would be wrong as none of them is better than another. Their usage should depend upon the objectives of any particular study. In a situation, one could be better whereas in another case the other one. The relation between both is more complementary rather than competitive. As also found by Tenenhaus (2008) and cited in Hair, Ringle, & Sarstedt (2011)–When the study is using suitable measures and data or when CB-SEM assumptions were met then in that situation both the approaches practically yield the same results. That is if before the application of PLS-SEM if measurement model characteristics are checked, then it will give similar results as of CB-SEM (Hair, Ringle, & Sarstedt, 2011).

If precisely discussing the strength of PLS, then it is statistically more robust, sensitive, and simpler with less strict assumptions. However, its inability to deal with nonrecursive relationships and causal loops; and the absence of global goodness of fit acts as its weakness.

In the dynamic business environment, VUCA is quite high, and almost every data in the social sciences follows Pareto's principle. It does not showcase normality and is mainly skewed, also as the life cycle of every product and service is shortening due to technological advancements, it becomes imperative to finish the marketing research very rapidly. Therefore, it becomes difficult to fulfil all the assumptions of the CB-SEM and to have a large sample size, which acts as an opportunity for the usage of PLS. Seeing the dynamic nature of the ever-changing business scenario, this methodology facilitates the researcher with appropriate judgmental decision choices.

However, every researcher should first identify her or his objectives and then decide the better-suited method for her or his study. Dijkstra & Henseler (2015) gave consistent PLS (PLSc) model, which is an extension of PLS and is comparable to covariance-based SEM, in future, the researchers suggest a study differentiating PLSc and CB-SEM. Researchers believe that PLSc-Sem usage has been overly used or indeed misused due to its simplicity or they believe that it will require less pain due to lesser strict assumptions. Whereas, they must be considering their objectives as their selection criterion to choose any method.

**References:**


