The Use of Partial Least Squares Path Modeling and Generalized Structured Component Analysis in International Business Research: A Literature Review

Joshua D. Shackman
Trident University

Structural equation modeling (SEM) is by far the best known and most widely used path modeling technique in the international business literature. However, recently international business researchers have begun to use a lesser known path modeling technique called partial least squares (PLS). PLS offers some advantages over SEM such as lower sample size requirements, easier testing of moderating relationships, and built-in capability to handle formative indicators which may explain the increased use by international business researchers. We examine the use of PLS in the international business literature, and the potential of a new path modeling technique called generalized structured component analysis (GeSCA) for international business research. We find mixed support for some of the commonly cited reasons for using PLS over SEM in the international business literature, but do find support for the use of PLS when sample sizes are not large enough for the use of SEM. Finally, we discuss GeSCA’s ability to handle multi-group analysis which may make it an attractive alternative over PLS for international business researchers who are using small sample sizes with data from multiple countries.

As pointed out in Hult et al. (2006), structural equation modeling (SEM) has become increasingly popular in the international business literature. The use of structural equation modeling has many well known advantages over other techniques such as standard regression analysis, as it allows for estimation and evaluation of an entire conceptual model rather than mere testing of individual hypotheses. Other advantages include the capability of estimating measurement error. One particular advantage for SEM in international business literature pointed out by Hult, et al. (2006) is the ability to compare models across groups, a useful feature for research that involves samples from multiple countries or cultures.

A lesser known path modeling technique is partial least squares (PLS). A method similar to SEM, Goodhue et al. (2006) find PLS to be widely used in the information systems but used only sparingly in other disciplines within management. However, a recent review by Henseler, et al. (2009) finds a growing use of PLS in the international marketing literature including several marketing studies published in top tier international business journals. PLS offers several potential advantages to international business researchers such as the smaller sample size requirements for SEM compared to PLS, and a lack of distributional assumptions. However, a shortcoming of PLS is that it does not provide a global fit statistic for models. This is a particular limitation for international business researchers in that multiple group comparisons become more difficult when an overall fit statistic is not available.
A third path modeling technique is called generalized structured component analysis (GeSCA). GeSCA may be of particular interest to international business researchers in that it combines both the small sample capabilities of PLS but like SEM also computes an overall fit statistic. GeSCA software allows for multiple group analysis similar to those computed in SEM software. There is also some preliminary research that suggests that under some circumstances GeSCA may provide more accurate parameter estimates than PLS. However, to date this technique has yet to have been used in a published international business study.

In this paper we first give a brief overview of the use of PLS in the international business literature. We focus on the reasons cited for using PLS over SEM in ten recent studies that were published in top ranking international business journals since 2001. We discuss the use of multiple group analysis, a technique well suited to international business research, in PLS, SEM, and in GeSCA. We then examine the validity of the reasons chosen for using PLS over SEM based on our review of the Monte Carlo simulation research, and also discuss some preliminary Monte Carlo simulation research using GeSCA with implication to international business researchers.

Advantages of PLS Cited in the International Business Literature

Numerous review articles have compared PLS and SEM, particularly within the discipline of information systems where the relative merits of each method have been hotly debated (Chin, 1995; Chin, 1998; Marcoulides & Saunders, 2006). The reader is referred to these reviews for a more technical discussion. In this section we will focus on just a few key practical differences that might affect an international business researcher’s decision to use PLS or SEM. We focus on the reasons for using PLS in ten recently published international business studies in the Journal of International Business, Management International Review, and the Journal of World Business. Table I summarizes the reasons given in each of these studies for their choice of PLS.

To begin with, one driving factor in the decision to use PLS or SEM is sample size. SEM is known to typically require samples of at least 200 with even larger samples required for more complex models (Kline, 2010). A failure to provide a large enough sample can result in a lack of convergence to a solution. By comparison, PLS uses a different algorithm to compute solutions involving principal component analysis rather than maximum likelihood factor analysis that allows solutions to be reached with far smaller sample sizes. Of the ten studies we reviewed, five cited small sample sizes as part of the rationale for choosing PLS. The sample sizes in the ten studies we reviewed ranged from 51 to 274. Some of these studies used very narrow and specific populations where a large sample size may have been difficult to obtain. For example, Boehe (2010) uses a sample of managers of product engineering departments of multi-national firms located in Brazil.

Another commonly cited factor in the decision to use PLS or SEM involves the distributional assumptions required. SEM assumes a normal distribution for the data, whereas PLS requires no distributional assumptions. Of the ten studies we reviewed,
seven of these studies cited lack of distributional assumptions as a reason to use PLS. Computations in SEM assume that the data meets certain normal distribution requirements but this is not an assumption for the computation of the parameters in PLS. Also, PLS is cited as being more appropriate for exploratory research where theory is at an early stage of development. SEM is considered to be more of a confirmatory technique when theory is at a relatively advanced stage of development. Exploratory research was a reason cited for using PLS in five of the ten studies we reviewed.

Another reason commonly cited for using PLS is the type of measurement scale being used in the study. PLS software such as SmartPLS have built in for the use of both reflective and formative scales, whereas SEM software is typically designed mostly for reflective scales. Some SEM software such as LISREL can handle formative scales, but it is believed to be far more complicated to include formative scales in LISREL than it is to include them using PLS (Venaik, et al., 2004). Five of the ten studies we reviewed cited the use of formative scales as a rationale for using PLS, and a sixth study (Acedo & Jones, 2007) uses a formative scale even though this is not a reason they cite for using PLS. The reader is referred to more extensive reviews for a more detailed discussion of the difference between formative and reflective scales (Mackenzie, et al., 2005; Coltman, et al., 2008).

One of the key differences between PLS and SEM not discussed in any of the ten studies we reviewed is that SEM computes several fit statistics that can be used to evaluate the overall fit of the model. Such fit statistics are not computed in PLS. Instead, the model must be evaluated based on the r-squareds of the different dependent and mediating variables in the model making comparison between different models more difficult. This also limits the ability of PLS to perform multiple group analysis, an important consideration for international business researchers discussed in the next section.

Multiple Group Analysis

One important method in path modeling for international business researchers mentioned both Hult et al.’s (2006) review of the use of SEM and Henseler’s (2009) review of PLS is multiple group analysis. As Hult et al. (2006, page 407) point out “because IB research often crosses multiple boundaries such as language, culture, politics, and economics, there is an increased need for measurement equivalence that allows for valid and accurate comparison of multiple groups”. Multiple group analysis is a technique built into SEM software such as AMOS that allow a researcher to readily compare model parameters across different samples such as countries or cultures.

A shortcoming of PLS is that none of the current PLS software packages have built in multiple group analysis capabilities. At least two of the ten studies summarized in Table I use multi-country samples that could benefit from multiple group analysis. For example, Fey et al. (2009) uses data from three different countries – Russia, the U.S., and Finland. Lee, et al. (2006) uses data from MBA students in the U.S. and Hong Kong. The multiple group analysis approach used in SEM requires a fit statistic, but PLS does not compute a fit statistic so different methods much be used. Henseler (2009) propose a
multiple group analysis approach in PLS using bootstrap methods to compare parameters between groups, but this technique is not yet built into any current PLS software package.

International business researchers wishing to conduct a multiple group analysis but do not have a sample size large enough to use SEM may wish to consider a third path modeling technique called generalized structured component analysis (GeSCA) (Hwang & Takane, 2004). GeSCA is similar to PLS but uses a slightly different algorithm that computes an overall fit statistic similar to the ones computed in SEM. Like PLS, GeSCA requires relatively small sample sizes to compute solutions and like PLS can easily accommodate

**Table I: Reasons Given for the Use of PLS in Published International Business Studies**

<table>
<thead>
<tr>
<th>Exploratory nature of study</th>
<th>Small sample</th>
<th>Lack of distributional assumptions</th>
<th>Formative indicators</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acedo and Jones (2007, JWB)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>216</td>
</tr>
<tr>
<td>Ainuddin, Beamish, Hulland, and Rouse (2007, JWB)</td>
<td>X</td>
<td></td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>Boehe (2010, MIR)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>146</td>
</tr>
<tr>
<td>Fey, Morgulis-Yakushev, Park, and Bjorkman (2009, JIBS)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>241</td>
</tr>
<tr>
<td>Lee, Yang, and Graham (2006, JIBS)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>176</td>
</tr>
<tr>
<td>Navarro, Losado, Ruiz, and Di’ez (2010, JWB)</td>
<td>X</td>
<td></td>
<td>X</td>
<td>150</td>
</tr>
<tr>
<td>Shi, White, Zou, and Cavusgil (JIBS, 2010)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>274</td>
</tr>
<tr>
<td>Venaik, Midgley, and Devinney (2005, JIBS)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>163</td>
</tr>
<tr>
<td>Venaik, Midgley, and Devinney (2004, MIR)</td>
<td>X</td>
<td></td>
<td></td>
<td>165</td>
</tr>
<tr>
<td>West and Graham (2004, MIR)</td>
<td>X</td>
<td></td>
<td></td>
<td>51</td>
</tr>
</tbody>
</table>

formative scales. The capability to compute a fit statistic makes multiple group analysis relatively straightforward in GeSCA, and software has been developed with built-in capability to use multiple group analysis using GeSCA. This is a very recently developed software package that has not yet been widely used in the management literature. The next section discusses some of the Monte Carlo simulation research that compares the accuracy of SEM, PLS, and GeSCA.

**Comparison of SEM, PLS, and GeSCA**

Monte Carlo simulation research has generally confirmed the appropriateness of the use of PLS when samples are relatively small. Chin and Newsted (1999) find that PLS provides more accurate parameter estimates than regression analysis when sample sizes are small, a result confirmed by Goodhue et al. (2006). Results are more mixed when comparing PLS and SEM in small samples. Goodhue et al. (2006) find SEM to provide more accurate parameter estimates than PLS even when sample sizes are small when using a Monte Carlo simulation with a simple model. Using a Monte Carlo simulation with a relatively more complex model, Reinartz, et al. (2009) find that PLS provides more accurate results when sample sizes are below 250. The sample sizes in the ten studies we reviewed have sample sizes from 51 to 274, similar to the range of sample sizes found in Reinartz, et al. (2009) to be appropriate for the use of PLS and perhaps too small to allow SEM to converge to a solution.

While the lack of distributional assumptions is cited in seven of the ten studies we reviewed, Monte Carlo simulation does not support this as a justification to use PLS over SEM. Reinartz, et al. (2009) find SEM to be highly robust to different distributions. No advantage of PLS over SEM is found for non-normally distributed data. Similar results are found in simulation research by Hsu, et al. (2006). Of all of the justifications cited in for the use of PLS over SEM, the lack of normally distributed data appears to have the least support.

Results are more mixed regarding the use of PLS for exploratory based research. Reinartz, et al. (2009) find that PLS generally has greater statistical power for detecting statistically significant relationships within a model (e.g. lower probability of a Type II error). They argue that the ability to detect significant relationships within the data makes it more suited for exploratory research and theory building. Goodhue et al. (2006) however does not find evidence of greater statistical power with PLS over SEM. The difference in results may be due to the simpler model used in the Goodhue, et al. (2006) simulation compared to the relatively more complex model used in Reinartz, et al. (2009).

A recent Monte Carlo simulation study comparing GeSCA to PLS and SEM finds that GeSCA provides more accurate parameter estimates than PLS when models are correctly specified (Hwang, et al, 2010). Under conditions of model misspecification, it is found that GeSCA provides more accurate parameter estimates than SEM (but similar estimates to PLS). The results regarding model misspecification lend credence to using PLS or GeSCA when the theory is at an early stage of development and one has less confidence in the precise specification of the model.
Conclusions

While we have only found ten studies using PLS that were published in the top international business journals over the last ten years, six of them have been published since 2007. This is similar to Hult, et al.’s (2006) finding of a rapidly growing use of SEM in the international business literature and Henseler, et al.’s (2009) finding of a rapidly growing use of PLS in the international marketing literature. This rising use of both PLS and SEM suggests international business researchers should familiarize themselves with the options they have available to them in terms of techniques and software packages available.

Of the five reasons cited for the decision to use PLS summarized in Table I, small sample size appears to be the most convincing reason. Only three of the studies we reviewed had sample sizes over 200, and none had a sample size over 300. While sample size was mentioned as a reason for use of PLS in five of these ten studies, it was likely a factor for all of these studies given the sample size requirements of SEM. The sample populations in these ten studies tended to be relatively specialized and narrow, which may have limited the researchers’ ability to obtain larger sample sizes. The choice for these researchers may not have been one between PLS and SEM given the relatively small sample sizes, rather the choice may have been between SEM and multiple regression. Monte Carlo simulations generally show that PLS provides more accurate results than regression, so the use of PLS over regression appears to be well supported given the size of the samples in the studies we reviewed.

Multiple group analysis is mentioned as a significant concern for international business researchers both when using SEM and PLS. However, we did not see the use of multiple group analysis in any of the ten studies we reviewed. Hult et al. (2006) also finds that multiple group analysis is not widely used within SEM in the international business literature. This may be due to the complexity or sample size requirements in SEM, and also due to the lack of built-in multiple group analysis features in PLS software. For this reason, we find the development of GeSCA to be an intriguing option for international business researchers given its built in ability to perform multiple group analysis with relatively small sample sizes. Little if any research to date has been done on the comparison of GeSCA, PLS, and SEM when using multiple group analysis so this is clearly a promising area for future research and one area to watch for international business researchers.

References


Contact email address: joshua.shackman@trident.edu