A New and Simple Approach to Multi-Group Analysis in Partial Least Squares Path Modeling

Jörg Henseler, Nijmegen School of Management, Radboud University Nijmegen,
Thomas van Aquinostraat 1, 6525 GD Nijmegen, The Netherlands,
Phone: +31 24 361 1854, Fax: +31 24 361 1933, E-mail: j.henseler@fm.ru.nl

Introduction: The Need for Multi-Group Analysis in PLS Path Modeling

In many sciences, and in particular in business and social sciences, partial least squares path modeling has become an established means for modeling complex relationships between latent variables. Very often, researchers face a heterogeneity of observations, i. e. for different sub-populations, different population parameters hold. For instance, institutions releasing national customer satisfaction indices may want to know whether model parameters differ significantly between different industries (c. f. Fornell 1992). Another example are customer segmentation studies based on FIMIX-PLS (c. f. Ringle, Wende, Will 2005). Usually, the segmentation phase is followed by separate PLS path modeling analyses of the customer segments, whose outcomes are subject to inter-segment comparisons. Commonly in multi-group analysis, a population parameter β is hypothesized to differ for two subpopulations, i. e. $\beta_{(1)} \neq \beta_{(2)}$. Referring to PLS path modeling, researchers would ask whether differences in path coefficients between subsamples, say $b_{(1)}$ and $b_{(2)}$ are significant.

An Overview over Existing Approaches to Multi-Group Analysis in PLS Path Modeling

Basically, three approaches to multi-group analysis in PLS path modeling have been proposed in literature so far: the parametric approach, the moderation approach, and the permutation approach.

The parametric approach. Keil et al. (2000) based their statistical test of group differences on the pooled standard errors that they had gathered from Jackknifing. Similarly, Chin (2000) suggests as a quick fix to treat the estimates of the bootstrap re-sampling in a parametric sense via t-tests. After having exposed the subsamples to separate bootstrap analyses and having made parametric assumptions about the distributions of the parameter standard errors, one may calculate the following statistic for the difference in paths between groups. It is asymptotically *t*-distributed with $n_{(1)} + n_{(2)} - 2$ degrees of freedom.

$$t = \frac{b_{(1)} - b_{(2)}}{\sqrt{\frac{(n_{(1)}-1)^2}{n_{(1)}+n_{(2)}-2} \cdot se^2_{\beta_{(1)}} + \frac{(n_{(2)}-1)^2}{n_{(1)}+n_{(2)}-2} \cdot se^2_{\beta_{(2)}}} \cdot \sqrt{\frac{1}{n_{(1)}} + \frac{1}{n_{(2)}}}} \qquad \text{Equation 1}$$

The subsample-specific path coefficients are denoted as *b*, the sizes of the subsamples as *n*, and the path coefficient standard errors as resulting from bootstrapping as *se*.

The moderation approach. Baron and Kenny (1986) define moderator variables as metric or categorical variables that influence the strength and/or direction of a relationship between two other variables. In the terminology of Baron and Kenny, group effects are nothing else than moderating effects of a categorical moderator variable, namely the grouping variable. Henseler and Fassott (forthcoming) as well as Tenenhaus (forthcoming) advocate to use Chin, Marcolin & Newsted's (2003) well-known interaction approach for PLS path modeling in order to test for group effects. The moderation approach uses bootstrapping to test the group effect hypothesis and hence does not rely on distributional assumptions like the parametric approach. However, due to the standardization of latent variable scores within the PLS path modeling algorithm (c.f. Tenenhaus et al. 2005), it is not possible to determine the size of a potential group effect.

The permutation approach. The permutation approach was coined by Chin (2003) and illustrated by Chin and Dibbern (forthcoming). In analogy to Edgington (1987), the procedure of the PLS-based permutation test is carried out as follows:

1. A test statistic is computed for the data, for instance the difference between two path coefficients.
2. The data is repeatedly permuted (maintaining consistency with the random assignment procedure).
3. The test statistic is also calculated for each permutation.
4. The proportion of the test statistics resulting from the permutation that exceeds/falls below the value stemming from the original data, determine the error probability. If for instance more than 95% of the permutation test statistics exceed the original test statistic, the null hypothesis should be rejected.

The permutation approach overcomes both the disadvantages of the parametric approach and the moderation approach. However, the central drawback of the permutation approach is that it is currently not available to PLS users, because it is not implemented in any known PLS software so far.

A New Approach to Multi-Group Analysis in PLS Path Modeling

From a procedural perspective, the new approach most resembles the parametric approach. Firstly, the subsamples are exposed to separate bootstrap analyses, and the bootstrap

outcomes serve as basis for the hypothesis tests of group differences. However, in contrast to the parametric approach, the new approach does not need any distributional assumptions. Instead, it evaluates the observed distribution of the bootstrap outcomes. Given two subsamples with parameter estimates (e. g., a path coefficient), $b_{(1)}$ and $b_{(2)}$, the probability $P(\beta_{(1)} > \beta_{(2)})$ is to be determined. More often than not, in PLS path modeling, continuous distributions of $\beta_{(1)}$ and $\beta_{(2)}$ are not available. Instead, by means of bootstrapping, empirical cumulative distribution functions (CDF's), say F, are determined through the bootstrap parameter estimates b_j. In case of J bootstrap samples, the empirical CDF is:

$$F(x) = P(\beta \leqslant x) = 1 - \frac{1}{J} \cdot \sum_{j=1}^{J} \Theta(b_j - x) \qquad \text{Equation 2}$$

Here, Θ denotes the unit step function, which has a value of one, if its argument exceeds zero, and zero otherwise. The probability of having group differences in the population is thus:

$$P(\beta_{(1)} > \beta_{(2)}) = 1 - P(\beta_{(1)} \leqslant \beta_{(2)}) = 1 - F_{(1)}(\beta_{(2)}) \qquad \text{Equation 3}$$

At this stage, one can again make use of the characteristic of the bootstrap, namely that it delivers an empirical cumulative distribution function of the population parameter.

$$F_{(1)}(\beta_{(2)}) = \sum_{i=1}^{J} \frac{1}{J} \cdot F_{(1)}(b_{(2)i}) = 1 - \frac{1}{J^2} \sum_{i=1}^{J} \sum_{j=1}^{J} \Theta(b_{(1)j} - b_{(2)i}) \qquad \text{Equation 4}$$

Knowing the parameter estimates from bootstrapping for two subsamples and with the following formula at hand, researchers can easily verify how probable a difference in parameters between two subpopulations is, and hence test their hypothesis.

$$P(\beta_{(1)} > \beta_{(2)}) = \frac{1}{J^2} \sum_{i=1}^{J} \sum_{j=1}^{J} \Theta(b_{(1)j} - b_{(2)i}) \qquad \text{Equation 5}$$

This formula states that J^2 (i. e., all possible) comparisons of bootstrap parameters have to be made. The number of cases in which the parameter of the one group exceeds the parameter of the other group has to be determined and divided by the number of comparisons.

Conclusions

The new approach combines the advantages of the three approaches presented before. It is simple to use, because it relies on nothing more than the bootstrap outputs that are generated by the prevailing PLS implementations (e. g., PLS-Graph, SPAD-PLS, and SmartPLS), it does not affect the estimate of the group difference, nor does it require distributional assumptions. The way of determining the probability that a population parameter differs across two sub-populations is unique to the new approach. It uses the empirical cumulative

distribution functions provided by bootstrap re-sampling as the basis for calculating the probability of differences in subpopulation parameters. Researchers can easily conduct the final calculations with available spreadsheet software like for instance MS Excel.

References

Baron & Kenny (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research. Journal of Personality and Social Psychology, 51, 1173-1182.

Chin (2000). Frequently Asked Questions – Partial Least Squares & PLS-Graph. Available from: http://disc-nt.cba.uh.edu/chin/plsfaq.htm (accessed 21-02-2007).

Chin (2003). A Permutation Based Procedure for Multi-Group Comparison of PLS Models, in: Vilares, Tenenhaus, Coelho, Vinzi, Morineau (eds.) *PLS and Related Methods.* Proceedings of the PLS'03 International Symposium, Lisbon, pp. 33–43.

Chin & Dibbern (forthcoming). A Permutation Based Procedure for Multi-Group Analysis: Results of Tests of Differences on Simulated Data, in: Vinzi, Chin, Henseler, Wang (eds): Handbook of Partial Least Squares, Heidelberg, Springer.

Chin, Marcolin & Newsted (2003). A Partial Least Square Latent Variable Modeling Approach for Measuring Interaction Effects: Result from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study. *Information Systems Research* 14(2): 189–217.

Edgington (1987). *Randomization Tests*. New York, Marcel Dekker, 2nd ed.

Henseler & Fassott (forthcoming). Testing Moderating Effects with PLS Path Modeling. An Overview over the Available Approaches, in: Vinzi, Chin, Henseler, Wang (eds): Handbook of Partial Least Squares, Heidelberg, Springer.

Keil, Tan, Wei, Saarinen, Tuunainen & Wassenaar (2000). A Cross-Cultural Study on Escalation of Commitment Behavior in Software Projects. *MIS Quarterly* 24(2), 299–325.

Ringle, Wende & Will (2005). Customer Segmentation with FIMIX-PLS, in: Aluja, Vinzi, Casanovas, Morineau & Tenenhaus (Ed.) *PLS and Related Methods.* Proceedings of the PLS'05 International Symposium, Barcelona, pp. 507–514.

Tenenhaus (forthcoming). A ULS/PLS Approach to Moderating Effects, in: Vinzi, Chin, Henseler, Wang (eds): Handbook of Partial Least Squares, Heidelberg, Springer.

Tenenhaus, Vinzi, Chatelin, Lauro (2005). PLS Path Modeling. *Computational Statistics and Data Analysis* 48(1), pp. 159–205.