PLS Path Modeling in Marketing and Genetic Algorithm Segmentation

Christian M. Ringle, University of Hamburg and University of Technology Sydney, cringle@econ.uni-hamburg.de
Marko Sarstedt, Ludwig-Maximilians University Munich, sarstedt@bwl.lmu.de
Rainer Schlittgen, University of Hamburg, rainer.schlittgen@uni-hamburg.de

Abstract

This paper presents the PLS genetic algorithm segmentation methodology which uses directed random searches to detect an optimum solution in the complex search space that underlies data partitioning tasks in PLS path modeling. The results of a simulation study allow a primary assessment of this novel approach and reveal its capabilities and effectiveness. Furthermore, applying the approach to the American Customer Satisfaction Index model allows unobserved heterogeneity and different consumer segments to be uncovered.

Keywords: Partial least squares, segmentation, latent class, ACSI, genetic algorithms

PLS Path Modeling in Marketing and Genetic Algorithm Segmentation

Introduction

Segmentation is a critical issue for marketing (e.g., Wedel and Kamakura, 2000) and partial least squares (PLS) path modeling (Lohmöller, 1989) applications in this field. The assumption that the data have been collected from a single homogeneous population is often unrealistic. Group-specific PLS path model estimates can significantly differ from those of other groups or the overall model (Ringle, Sarstedt and Mooi, 2009). Consequently, the failure to account for this heterogeneity may easily result in misleading interpretations (Jedidi, Jagpal and DeSarbo, 1997). On the manifest data level, sequential segmentation strategies, such as k-means or tree clustering, usually fail to identify groups of data with distinctive inner path model estimates (Sarstedt and Ringle, 2010). Researchers have addressed these problems by proposing novel segmentation approaches such as FIMIX-PLS (Hahn et al, 2002). Sarstedt (2008) evaluates these techniques and shows that they still suffer from deficiencies which relate to, for example, the types of heterogeneity covered or distributional assumptions.

The authors of this paper pursue three important objectives in contributing to data segmentation in PLS path modeling. First, we introduce a new kind of PLS segmentation approach that uses a genetic algorithm (GA) to account for heterogeneity in the estimates for inner and outer path model relationships. The resulting PLS genetic algorithm segmentation method (PLS-GAS) has been designed to overcome the shortcomings of prior approaches in that it does not rely on distributional assumptions. Consequently, PLS-GAS fits the PLS method's non-parametric character, allows for the integration of formative measurement models, uncovers highly unbalanced segments, and is not affected by extremely non-normal data. Second, we present the results of a simulation study that assesses the capabilities of PLS-GAS. Third, we test the usefulness of PLS-GAS in respect of one of the best known PLS applications in marketing literature: the American Customer Satisfaction Index (ACSI; Fornell et al., 1996).

In the next section, we provide a brief description of the PLS-GAS approach, followed by the simulation study and the empirical application. The paper concludes with a summary, limitations as well as future research directions.

Genetic Algorithm-Based Segmentation Methodology for PLS Path Modeling

When segmentation tasks are carried out in PLS path modeling, cases are assigned to a pre-determined number of clusters to uncover group-specific inner and outer path model relationships. Even small or midsized problems easily reach extremely high numbers of combinatorial solutions (Ringle and Schlittgen, 2007). For example the assignment of 300 observations into 2 groups entails about 10^{90} different outcomes. It is therefore normally impractical attempting to search for all possible assignments to uncover the best set of group-specific PLS estimates. To handle this complexity, the novel procedure applies the principles of evolution and natural genetics by using a GA-based clustering approach (Maulik and Bandyopadhyay, 2000), which is a randomized technique to search in large and multimodal landscapes. Several studies in different research contexts that entail NP-complete optimization problems, such as data clustering and job shop scheduling (Goldberg, 1975), have shown that GAs provide solutions which are close to or match the global optimum for an optimization problem's fitness (objective) function.

The PLS-GAS approach is a two-stage genetic/hill-climbing clustering hybrid (Cowgill, Harvey and Watson, 1999). In the first stage, a non-deterministic genetic algorithm is used to find the best possible starting partition by strolling through the search space and thereby approaching many local optima. This is done by minimizing a fitness criterion which considers the inner and outer models' residuals (Ringle and Schlittgen, 2007). As there is no guarantee that the GA outcome cannot be further improved, the second stage of PLS-GAS uses the best partition that was found for a deterministic hill-climbing approach to improve (if possible) the local partition for the ultimate best segmentation.
Each PLS-GAS run uses a fixed number of segments. The best fitting number of segments is usually unknown priori. Ringle, Sarstedt and Schlittgen (2009) proposed a two-stage approach in which FIMIX-PLS is first used on the data. Based on model selection statistics, the researcher decides on the segment number and uses this as input for the PLS-GAS analysis.

Simulation Study

To assess the performance of PLS-GAS, we use a simple path model (Figure 1) with two exogenous constructs (ξ_1 and ξ_2) and one endogenous construct (η_1) using a two-segment solution. The number of manifest variables is identical in the three outer models. As the analysis focuses on the inner model path relationships, we chose equal loadings for all outer models. The pre-specification of the two inner path model weights for data simulation uses a higher ω_{11} and a lower ω_{21} value for the first group of data and vice versa for the second group. As the sizes of the coefficients themselves are not important but their distinctiveness is, we focus on their alternative levels of difference in the inner model.

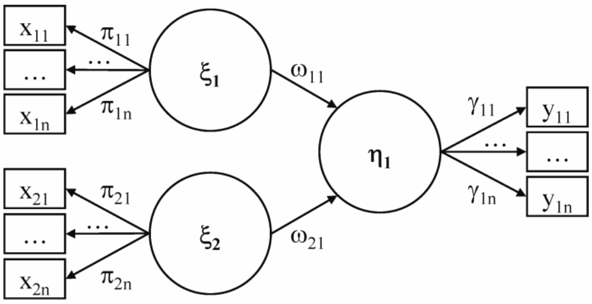
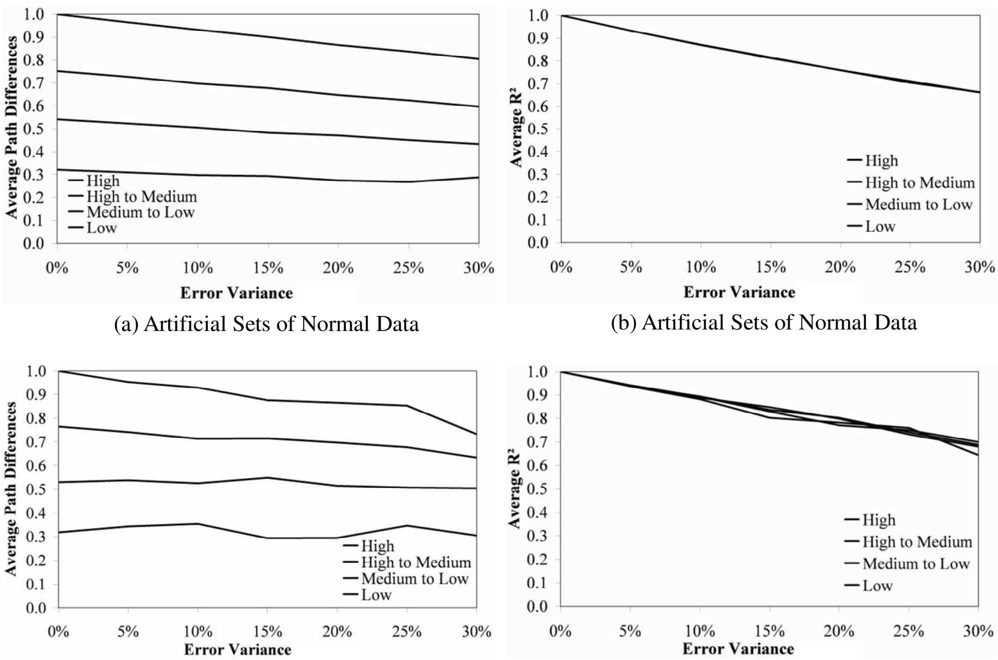


Figure 1: PLS Path Model

Five data characteristics are manipulated. The factors and their level are chosen on the rationale presented in previous simulation studies on covariance-based structural equations modeling (Boomsma and Hoogland, 2001) and PLS segmentation studies (Esposito Vinzi et al., 2007): (a) number of observations [100, 200, 600]; (b) identical number of manifest variables in the latent constructs' measurement models [3, 6, 12]; (c) size of the smaller segment [10%, 30%, 50%]; (d) differences between the group-specific inner model weights [high, high to medium, medium to low, low]; (e) error variance σ^2 of the endogenous latent variable as well as the manifest variables in relation to their respective total variance [0%, 10%, 20%, 30%]. The design is factorial; with two replications (datasets) per cell, we generate a total of $3^3 4^2$=432 experimental datasets for this study. See the study by Chin, Marcolin and Newsted (2003) for more details on the generation of datasets.
A first important investigation into the results addresses PLS-GAS's potential to meet the global optimum solution. The results of this analysis reveal that PLS-GAS provides a 100% correct assignment in all σ^2=0 constellations which provides evidence that PLS-GAS consistently achieves the global optimum solution in data constellations not affected by error

variance. Consequently, this method permits a clear-cut assignment of the experimental data sets. In addition, (a) neither the number of observations, (b) nor the number of manifest variables per measurement model has a significant effect on PLS-GAS computational results. Furthermore, PLS-GAS identifies the segments in accordance with their expected size in (c) all systematically changed relative segment size constellations. The excellent segmentation characteristics of this methodology are highly beneficial in practical applications that, for example, usually deal with unbalanced groups of observations. The only two factors that reduce the quality of parameter recovery relate to (d) a decrease in group-specific PLS path coefficients' distinctiveness and/or (e) an increased level of σ^2. This is shown in Figure 2, which incorporates the constellation with 100 observations per segment, six manifest variables per measurement model, as well as alternative inner model weight differences and levels of σ^2. In accordance with our previous findings, these illustrations of the outcomes are representative of all constellations in our analytical design, since (a) the sample size, (b) the numbers of indicators per measurement model, and (c) the relative segment size do not involve significant changes in the PLS-GAS segmentation outcomes. Whereas Figure 2 (a) shows the average path differences for various levels of error variance based on artificially generated raw data, Figure 2 (c) illustrates the PLS-GAS analysis results for these constellations. Similarly, Figures 2 (b) and (d) report average R^2 values.



(a) Artificial Sets of Normal Data (b) Artificial Sets of Normal Data

(c) PLS-GAS and Normal Data (d) PLS-GAS and Normal Data

Figure 2: Average Path Differences and R^2 Values (Sample Size 100/100)

The finally formed PLS-GAS segments (Figure 2 (c)) exhibit group-specific path differences that consistently match the levels expected from the artificially formed data sets (Figure 2 (a)). Moreover, Figure 2 ((b) and (d)) illustrates PLS-GAS computations in respect of the endogenous construct's R^2 values as σ^2 increases. On average, all group-specific R^2 outcomes meet the assumptions regarding the a-priori generated sets of data almost exactly. Consequently, PLS-GAS offers the capability to constantly perform well in all situations of changed factor levels. Additional analyses using non-normal data and higher numbers of segments reveal identical results, underlining the approach's capabilities.

Application of PLS-GAS to the American Customer Satisfaction Index Model

Next, we apply PLS-GAS to the ACSI model by Fornell et al. (1996)[1] but slightly modified the set-up by omitting the "Customer Complaints" construct measured by a binary single item. Parameter estimation is carried out using the statistical software application SmartPLS 2.0 (Ringle, Wende and Will, 2005). A systematic evaluation (Henseler, Ringle and Sinkovics, 2009) on the aggregate data level shows that the parameter estimates of the ACSI model exhibit high degrees of reliability and validity. For example, the minimum value for composite reliability in the ACSI application is .822 across all reflectively measured constructs which is clearly above the commonly suggested threshold value of .60.
To assess whether unobserved heterogeneity significantly affects parameter estimates, we apply PLS-GAS to the data. When applying the method, the researcher has to specifically determine the number of segments to be retained from the data set. Instead of merely reverting to proxies, such as the solution's interpretability or acceptable R^2 values, we used FIMIX-PLS (Hahn et al., 2002) on the data to guide this decision. Unlike competing procedures, FIMIX-PLS allows information and classification criteria to be computed. All criteria (AIC, BIC, CAIC, Entropy) uniformly indicate that a two-segment solution is deemed appropriate. Further analyses to determine higher segment numbers clearly indicate that the two-segment solution is most appropriate in terms of uncovering differentiable latent segments. Furthermore, the overall R^2 values are considerably higher than in solutions with more segments. Table 1 presents the global model as well as segment-specific PLS-GAS analysis results.

	global	PLS-GAS		
		Segment 1	Segment 2	|diff|
Customer Expectations of Quality → Perceived Quality	.556***	.791***	.305***	.486***
Customer Expectations of Quality → Perceived Value	.072***	.214***	.000	.214***
Customer Expectations of Quality → Overall Customer Satisfaction	.021***	.152***	-.032***	.185***
Perceived Quality → Overall Customer Satisfaction	.557***	.505***	.548***	-.043
Perceived Quality → Perceived Value	.619***	.533***	.622***	-.089
Perceived Value → Overall Customer Satisfaction	.394***	.338***	.419***	-.081
Overall Customer Satisfaction → Customer Loyalty	.687***	.677***	.694***	-.018
Relative segment size	1.000	.549	.451	
R^2 Perceived Quality	.309	.626	.093	
R^2 Perceived Value	.439	.510	.387	
R^2 Overall Customer Satisfaction	.777	.819	.747	
R^2 Customer Loyalty	.471	.458	.482	

*** sig. at .01, ** sig. at .05, * sig. at .10; |diff| = absolute differences between path coefficients; permutation-based multi-group comparison test by Chin and Dibbern (2009)

Table 1: ACSI Segmentation Results

Segment-specific reliability analyses reveal that all constructs exhibit a high degree of internal consistency. The validity analysis shows that most of the first segment's endogenous constructs achieve considerably higher R^2 values than the global model, thus indicating an

[1] The data were provided by: Fornell, Claes. AMERICAN CUSTOMER SATISFACTION INDEX, 1999 [Computer file]. ICPSR04436-v1. Ann Arbor, MI: University of Michigan. Ross School of Business, National Quality Research Center/Reston, VA: Wirthlin Worldwide [producers], 1999. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2006-06-09. We would like to thank Claes Fornell and the ICPSR for making the data available.

increased model fit. With the exception of “Perceived Quality,” which has a very low R^2 value of .093, all constructs in the second segment lie at similar levels when compared to the aggregate data analysis results. The weighted sum of R^2 values across the two segments leads to a considerably higher model fit of up to 25%.

When comparing the global model with the results derived from PLS-GAS, one finds that the relative importance of the driver constructs differs quite substantially across the two segments. For example, the global model suggests that customers' expectations of quality primarily influences their perception of quality (.556) and only exerts a minor influence on perceived value (.072) and overall satisfaction (.021). Whereas similar findings are possible regarding the second PLS-GAS segment, the importance of the customer expectations construct is far more pronounced in the first segment. Here, customers' expectations prior to purchase strongly influence all subsequent constructs. Consequently, customer expectations do not only directly influence overall satisfaction, but likewise exert a pronounced indirect effect on the model's primary target variable via mediating constructs. This is reflected in the total effect of customers’ expectations of quality on overall satisfaction, which lies considerably higher in the first segment (.767) than in the second segment (.215) and the global model (.496). These results clearly suggest that the data are heterogeneous, which the PLS-GAS procedure reveals.

PLS multi-group analyses (Chin and Dibbern, 2009) provide evidence that all paths related to the customer expectations construct differ significantly across the two segments. This reflects the varying importance of “Customer Expectations of Quality” in respect of all subsequent constructs (most notably overall customer satisfaction).

Summary and Conclusions

This paper contributes to the need for effective segmentation means in PLS path modeling by developing and evaluating a novel PLS segmentation methodology which uses a GA to cope with previous segmentation procedures’ deficiencies. PLS path modeling applications in marketing do not usually address the critical heterogeneity issue (Sarstedt, 2008). Rather, they relate their results’ presentation and interpretation to the unrealistic assumption that the data stem from a homogenous population. However, uncovering heterogeneity on the aggregate data level and compliant segmentation are two key issues in PLS path modeling if findings and conclusions are to be complete and valid. Conventional clustering techniques such as k-means or tree clustering usually fail to identify groups of data with distinctive inner path model estimates (Sarstedt and Ringle, 2010). The contribution of this research is to present a novel PLS segmentation method which permits homogenous groups of observations to be formed that exhibit significantly distinct PLS estimates. Researchers and practitioners require this kind of PLS path modeling method to obtain further differentiated analytical outcomes. Findings, interpretations, and conclusions become more precise with each formed segment.

Notwithstanding the usefulness of the segmentation methodology to further differentiate results if heterogeneity significantly affects PLS estimates, this research has some limitations. The literature does not as yet offer a means with which to accurately generate data with pre-specified parameters for formative PLS path models. This kind of artificial data generation method is a key requirement to evaluate PLS-GAS's performance regarding formative outer PLS models in future research. Furthermore, the consideration of other factor levels (such as higher segment numbers) and the exploration of potentially compounding effects of specific data characteristics such as multicollinearity would be promising to further explore PLS-GAS’s capabilities.

References

Boomsma, A., Hoogland, J. J. 2001. The Robustness of LISREL Modeling Revisited. In: Cudeck, R., du Troit, S., Sörbom, D. (Eds.), Structural Equation Modeling: Present and Future. Scientific Software International, Chicago, pp. 139-168.

Chin, W. W., Dibbern, J. 2009. A Permutation Based Procedure for Multi-Group PLS Analysis: Results of Tests of Differences on Simulated Data and a Cross Cultural Analysis of the Sourcing of Information System Services between Germany and the USA. In: Esposito Vinzi, V., Chin, W. W., Henseler, J., Wang, H. (Eds.), handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields. Berlin: Springer, forthcoming.

Chin, W. W., Marcolin, B. L., Newsted, P. N. 2003. A Partial Least Squares Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an electronic Mail Emotion/Adoption Study. Information Systems Research 14 (2), 189-217.

Cowgill, M. C., Harvey, R. J., Watson, L. T. 1999. A Genetic Algorithm Approach to Cluster Analysis. Computers and Mathematics with Applications 39 (7), 99-108.

Esposito Vinzi, V., Ringle, C. M., Squillacciotti, S., Trinchera, L. 2007. Capturing and Treating Unobserved Heterogeneity in Response-Based Segmentation in PLS Path Modeling: A Comparison of Alternative Methods by computational Experiments. ESSEC Research Center, Working Paper No. 07019, Cergy Pontoise Cedex.

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., Bryant, B. E. 1996. The American Customer Satisfaction Index: Nature, Purpose, and Findings. Journal of Marketing 60 (4), 7-18.

Goldberg, D. E. 1975. Genetic Algorithms in Search, Optimization and Machine Learning, Boston: Kluwer.

Hahn, C., Johnson, M. D., Herrmann, A., Huber, F. 2002. Capturing Customer Heterogeneity Using a Finite Mixture PLS Approach. Schmalenbach Business Review 54 (3), 243-69.

Henseler, J., Ringle, C. M., Sinkovics, R. R. 2009. The Use of Partial Least Squares Path Modeling in International Marketing. In: Sinkovics, R. R., Ghauri, P. N. (Eds.), Advances in International Marketing. Bingley: Emerald, pp. 277-320.

Jedidi, K., Jagpal, H. S., DeSarbo, W. S. 1997. Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity. Marketing Science 16 (1), 39-59.

Lohmöller, J.-B. 1989. Latent Variable Path Modeling with Partial Least Squares, Physica, Heidelberg.

Maulik, U., Bandyopadhyay, S. 2000. Genetic Algorithm-based Clustering Technique. Pattern Recognition 33, 1455-65.

Ringle, C. M., Schlittgen, R. 2007. A Genetic Segmentation Approach for Uncovering and Separating Groups of Data in PLS Path Modeling. In: Martens, H., Næs, T., Martens, M. (Eds.), Causalities Explored by Indirect Observation: Proceedings of the 5th International Symposium on PLS and Related Methods (PLS'07). Åas, Norway: Matforsk.

Ringle, C. M., Sarstedt, M., Mooi, E. A. 2009. Response-Based Segmentation Using Finite Mixture Partial Least Squares: Theoretical Foundations and an Application to American Customer Satisfaction Index Data. Annals of Information Systems 8, forthcoming.

Ringle, C. M., Sarstedt, M., Schlittgen, R. 2009. Finite Mixture and Genetic Algorithm Segmentation in Partial Least Squares Path Modeling. In: Fink, A., Lausen, B., Seidel, W., Ultsch, A. (Eds.), Studies in Classification, Data Analysis, and Knowledge Organization, Proceedings of the 32nd Annual Conference of the German Classification Society. Heidelberg, Berlin: Springer, forthcoming.

Ringle, C. M., Wende, C., W, Will. 2005. SmartPLS 2.0. Available from www.smartpls.de, accessed 30 June 2009.

Sarstedt, M. 2008. A Review of Recent Approaches for Capturing Heterogeneity in Partial Least Squares Path Modelling. Journal of Modelling in Management 3 (2), 140-61.

Sarstedt, M., Ringle, C. M. 2010. Treating Unobserved Heterogeneity in PLS Path Modelling: A Comparison of FIMIX-PLS with Different Data Analysis Strategies. Journal of Applied Statistics, forthcoming.

Wedel, M., Kamakura, W. 2000. Market Segmentation: Conceptual and Methodological Foundations, 2nd edition. Kluwer Academic Publishers, Boston.

ISBN 1 86308 158 5

ANZMAC Website :
www.anzmac.org

Peer Reviewed Papers: The conference papers included on this website were double-blind, peer reviewed in accordance with DIISR requirements

Conference Secretariat
Promaco Conventions
PO Box 890, Canning Bridge WA 6153
Ph: (+61 08) 9332 2900 Fax: (+61 08) 9332 2911
www.promaco.com.au

Publisher: Australian & New Zealand Marketing Academy (ANZMAC)