Capturing Unobserved Heterogeneity in Structural Equation Models: from Maximum Likelihood to Partial Least Squares

Vincenzo Esposito Vinzi
ESSEC Business School, France
vinzi@essec.fr

In many companies, the evaluation of business performances in terms of customer satisfaction is crucial. In common practice customer satisfaction is meant as a latent variable and therefore indirectly measured as a linear combination of observed indicators. Furthermore, it is important to identify which factors mainly contribute to the increase or to the decrease of the customer satisfaction level and what are the consequences of this variation in terms of loyalty.
In this framework, PLS (Partial Least Squares) approach to structural equation modelling (Tenenhaus, Esposito Vinzi, Chatelin and Lauro, 2005) is largely used to estimate impacts between drivers of satisfaction, satisfaction itself and related consequences as well as individual scores for all these factors (Esposito Vinzi, Chin, Henseler and Wang, 2007).
When measuring customer satisfaction, a typical research issue is the identification of distinctive customer segments by capturing the so called unobserved heterogeneity of customers' behaviour. In the classical maximum likelihood approach to structural equation modelling, segmentation can be achieved based on finite mixture models and the expectation-maximization (EM) algorithm (Jedidi et al. 1997a, 1997b). In this approach to classification, the data is supposed to be the result of the mixture of two or more populations mixed in different proportions. In other words, each subject is supposed to belong to a segment, each segment being characterised by a different covariance structure. Hence, data arise from a mixture of distributions, and the aim is to estimate the probability that each subject belongs to each of these sub-populations. Distributions that are more frequently used are the multivariate normal or multinomial distribution. The technique is based on the EM algorithm, which allows the estimation of the posterior probabilities. The posterior probabilities represent a fuzzy classification of the observations in segments based on the postulated measurement and structural models. However, this solution seems to be inappropriate for PLS because of hard methodological assumptions that would be incoherent with the soft modelling PLS spirit.
Hahn et al. (2002) propose the finite mixture partial least squares (FIMIX-PLS) approach that is however strongly connected to the maximum likelihood theory. According to their proposal, a finite mixture procedure is joined with an EM-algorithm specifically regarding the latent variable scores predicted in PLS. Therefore, some conceptual and methodological problems arise and demand for further research. Among others, it is worth mentioning that the measurement model is assumed to be the same for all groups, the multivariate normality for the estimated latent variables is assumed, the EM algorithm might fall in local optima.
Moreover, an accurate model assessment is essential to ensure a statistical validation of the obtained results in terms of differences between yielded models (Tenenhaus et al., 2007).
Finally, the validated models shall undergo an ex post analysis aiming at a suitable characterization of the identified groups for further interpretation purposes and additional insights.

The most important fact within this context is that, in case heterogeneity of behaviours holds, poor standard PLS results for the overall dataset may turn into good estimates of the path coefficients and substantial values for R^2 of latent endogenous variables after segmentation. Therefore, it is extremely important to look for unknown group structure for which significantly different models exist.

Main References

Esposito Vinzi, V., Chin, W., Henseler, J., Wang, H. (Eds.), Handbook of Partial Least Squares (PLS): Concepts, Methods and Applications, Springer, (2007, to appear).

Hahn, C., Johnson, M., Herrmann, A., Huber, F., Capturing Customer Heterogeneity using a Finite Mixture PLS Approach, Schmalenbach Business Review, **54**, 243-269, (2002).

Jedidi, K., Harshanjeet, S. J., De Sarbo W.S., Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity, Marketing Science, **16**, 39-59, (1997a).

Jedidi, K., Harshanjeet, S. J., De Sarbo W.S., STEMM: A General Finite Mixture Structural Equation Model, Journal of Classification, **14**, 23-50, (1997b).

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.M., Lauro, C., PLS Path Modeling, Computational Statistics and Data Analysis, **48**, 159-205, (2005).

Tenenhaus, M., Mauger, E., Guinot, C., Use of ULS-SEM and PLS-SEM to measure interaction effect in a regression model relating two blocks of binary variables, Handbook of PLS and Marketing, Springer-Verlag, (2007, to appear).