Market Segmentation through Conjoint Analysis using Latent Class Models

Liberato Camilleri, Lara Marie Azzopardi
Department of Statistics and Operations Research
University of Malta
Msida (MSD 06) Malta
E-mail: liberato.camilleri@um.edu.mt

KEYWORDS
Conjoint Analysis, Market Segmentation, Latent Class Model, EM algorithm, Maximum likelihood estimation.

ABSTRACT

Conjoint Analysis is accepted by market researchers as a reliable and suitable instrument for measuring consumer preferences. The popularity of conjoint analysis hinges on the belief that it produces valid measurements of consumer preferences for the features of a product or service. It is the marketers' methodology for assessing the impact of proposed actions on the market and finding out how buyers trade-off among competing products and suppliers. A popular application of conjoint analysis is market segmentation which addresses the heterogeneity in consumer preferences. Market segmentation assumes that a heterogeneous population is represented as a collection of homogeneous subgroups where customers in each cluster have similar needs and similar views of how to worth a product. Other applications of conjoint analysis include pricing strategies, product positioning, competitive analysis, promotional policies, new product identification and distribution decisions. This paper describes the issues in implementing conjoint analysis and then illustrates the methodology to perform market segmentation using latent class analysis. The application focuses on customer preferences when evaluating the worth of mobile phones given demographic and product-related predictors.

1. IMPLEMENTING CONJOINT ANALYSIS

Conjoint analysis involves a framework of distinct steps, which include the selection of the utility (preference) function; selection of a method, design and procedure for data collection; the selection of a measurement scale for the response variable and the selection of an estimation method.

The utility function relates the benefit of a product profile to defined attributes (predictors). These attributes could either be discrete or continuous. There are basically three types of utility functions. The **vector model** assumes a linear relationship between the utility of a product and an attribute having a metric scale. The **ideal point model**, very often a quadratic function, assumes the existence of an ideal manifestation. The utility reaches a maximum value at one attribute value. The **part-worth model** relates the utility of a product to a categorical attribute. A parameter is estimated for each attribute category.

The four types of data collection methods for conjoint analysis include the self-explicated technique (Srinivasan and Wyner 1989); the full profile approach (Green and Rao 1971); the two-factor method (Johnson 1974) and the hybrid technique. **Self-explicated** techniques assess the subjects' utilities directly. The respondents are first asked to worth the levels of each attribute separately by rating them on a discrete preference scale and then asked to rate the importance of each attribute, perhaps using a different preference scale. Part-worths are computed by multiplying the importance weights with the attribute-level desirability ratings. In a **full profile** approach a respondent has to worth a complete set of profiles (stimulus cards) describing a product where each profile incorporates one level from each of the attributes of interest. The flexibility in scaling makes the full profile approach more attractive than other approaches. The main argument that favours the full profile approach is that it comes closer to a real buying situation in which the respondents react to a set of total profile descriptions, which are realistic representations of real items. In a **two-factor** approach, items are assessed through two-way attribute tables (trade-off matrices) in which the rows and the columns represent the levels of the two selected attributes. Each respondent has to rank all combinations of the levels of the two attributes in the associated matrix elements. The number of two-way tables, that has to be ranked, increases with the number of attributes used. An advantage of this approach is that it reduces information overload on respondents because all the attributes are evaluated two-at-time. The limitations of this approach are the exhaustive number of two-way tables that need to be filled out and the lack of realism in decomposing the set of attributes to two-at-time combinations. A **hybrid approach** combines the self-explicated task with aspects of the full profile conjoint analysis. The first part of the interview uses a self-explicated approach in which a respondent is asked to give a direct judgment of each attribute and its levels prior to the presentation of the profiles. The self-explicated context puts emphasis on evaluating products feature by feature rather judging the product as a whole. HCA (Hybrid conjoint analysis) uses full profiles in the second stage; whereas ACA (Adaptive conjoint analysis) uses partial profiles, composed of only a subset (usually two or three) of attributes, in paired-comparisons and which is viewed as a modern form of the two-factor method. The primary advantage of these hybrid techniques over other methods is that they allow a greater number of attributes to be managed by evaluating smaller numbers of profiles. The reduction in the number of profiles judged is compensated by the information collected from the self-explicated interview.

The three types of data collection designs include the complete factorial, the bridging and the fractional factorial designs. If a conjoint application is confined to a limited number of attributes with a limited number of levels then the full profile approach can be implemented by a **complete factorial design**. Such a design includes all possible combinations of the levels of the attributes in the study. This approach offers no problem with orthogonality (independence of the attributes) and all main effects and their interactions are estimable. The major limitation of this approach is that most applications include several attributes with varying number of levels. Implementing a complete factorial design would create a large number of incentives (item profiles) and will result in information overload on the respondents. The bridging and fractional factorial designs resolve this problem by reducing the number of incentives. In a **bridging design** the whole set of attributes is split into subsets and each card deck is composed of attribute level combinations from any subset of the attributes. To link part-worth functions across the various subsets of attributes one or two attributes will be common across all card decks. In a **fractional factorial design** the design is reduced systematically in such a way that the attributes are orthogonal as much as possible. In some commercial applications, the attributes are correlated and so an orthogonal design can produce stimuli that are not realistic. Other orthogonal displays can be tried by permuting sets of attribute levels if some of the stimulus profiles turn out to be non-representative.

Verbal, paragraph and pictorial descriptions are basically the three ways of presenting the incentives. In a **verbal** presentation the incentives are presented on information sheets using either key words or descriptive sentences or a combination of both. The **paragraph** description approach provides a more realistic and complete description of the stimuli and is used when comparing and testing different advertising claims. A drawback of this procedure is the information overload on respondents by having to read large quantities of information. Reducing the total number of descriptions may produce very inaccurate parameter estimates at the individual level. Another limitation is that verbal and paragraph descriptions are subject to response biases resulting from the order in which attributes are presented. The importance of an attribute is to some extent affected by the position of the attribute in the stimulus card. **Visual** presentations can either be graphic, where drawings or photographs are used or physical, where real products and prototypes are used. The use of profile cards and other pictorial material causes less fatigue to the respondents by providing an easier way to get information and hence allow a greater number of attributes to be included in the study. Another advantage is that the visual stimuli are more realistic because in the marketplace consumers choose their products by inspecting them. Such an inspection is more closely approximated by pictorial presentations. The use of film clips and full-scale prototypes is essential to give respondents maximum exposure to the stimulus especially when the task involves a radical new product idea. The primary disadvantage is that visual displays may exhibit additional information, such as style and colour of the item, that the researcher has no intention to analyze.

Three data collection procedures include person-to-person interviews or use mail and online questionnaires. Using **person to person** interviews is a rather slow process and very time consuming. The use of **mail questionnaires** ensures geographic representativeness but may suffer from lack of response. Phone-mail-phone procedures are used by several researchers to ensure a high completion rate with negligible missing data problems and simultaneously reduce the selection bias of respondents. A relatively new method for data collection is the **online questionnaire** in which the respondents receive the questionnaire and send their reply via e-mails.

The response modes used to evaluate incentives or stimuli can be divided into metric, non-metric and choice based. **Ranking** and **paired profile** comparisons are non-metric procedures. In rank data the outcome is just an order of preferences. It may express the preference-worthiness of a profile but does not result in metric ordinal preference data. In a paired-profile comparison the respondent has to declare his preference between two incentives. One of the shortcomings of rank-based data is the distortion caused by the interference between less and highly important variables. Rating, constant sum comparisons and dollar metrics are metric procedures. In **rating** data respondents grade the profiles subjectively on an interval scale, assuming that they perceive scale spacing. The outcome expresses the intensity of the preferences. In a **constant sum** comparison respondents are asked to allocate a fixed number of points across a number of profiles. This method provides importance weights that depend on the perceived importance of each profile. Another way of obtaining interval-scale judgments is the **dollar metric** approach. In this graded paired comparison a respondent has to compare two items and has to state the price that must be added to the least preferred item to make it equally worth to the other. The results are then aggregated to obtain an interval, scaled dollar metric of comparisons. Limitations to this approach are that respondents may have biased perceptions with regards to the use of price differences as a response measure and is a slow procedure compared to the rating method. **Choice-based** conjoint analysis relies on data from a discrete choice experiment in which each product is a hypothetical combination of attributes chosen by an experimental design procedure. The respondents are presented with profile descriptions of two or more competing items that vary on one or more attributes and their task is to choose the most preferred item. The major advantage of a choice-based task is that it has greater external validity because it mimics what consumers actually do in the marketplace. Moreover, it is a simpler task for respondents to choose incentives rather than rate or rank these alternatives. The major limitation of choice-based analysis is that it contains minimal information about consumer preferences. A choice simply indicates which profile is most preferred but it does not provide an estimate of the utility of the product profiles.

Modern statistical analysis is based on the likelihood principle that all the information in the observed data is contained in the likelihood. The likelihood can be defined as the probability of the observed responses expressed as a function of the unknown parameters. Hence a likelihood method can use the data optimally. Maximum likelihood and Bayesian analysis are the two main areas that use likelihood methods. Hierarchical Bayes methods derive part worths by combining information on the distribution across respondents. The posterior distribution of individual parameters is estimated using a computationally intensive method called Gibbs sampling that produces estimates of each respondent's part worths and standard errors. Hierarchical Bayesian analysis provides very flexible output and the researcher may choose among many possible population distributions; however the method requires considerable expertise to execute properly. During the last four decades researchers have used these estimation techniques to estimate parameters of models for different types of conjoint data.

2. MARKET SEGMENTATION

Traditionally, market segmentation in conjoint analysis was carried out using either a-priori or post hoc procedure. In a-priori segmentation analysis the number of segments is determined in advance by the researcher and individual-level preference judgments are combined at the segment level. Actually, this is not appropriate since demographic and psychographic predictors rarely describe adequately the heterogeneous utility functions. In post-hoc or tandem segmentation, estimation and clustering are carried out consecutively. Individual-level parameter estimates are first obtained from normal regression models and then individuals are clustered on the basis of similarity of the estimated parameters by using Ward's hierarchical or K-means non-hierarchical clustering procedures. This two-stage approach also has problems since different clustering methods will produce different outcomes. Moreover, the initial utility estimation method using regression analysis and the subsequent cluster analysis optimize different and unrelated objective functions.

To address the limitation of a-priori and post-hoc methods, several integrated conjoint segmentation methods were proposed in which the parameters within the segments are estimated at the same time that the segments are identified. Thus a single criterion of interest is optimized under a set of constraints. (Hagerty 1985) proposed a method using a weighting scheme representing a factor-type partitioning of the sample. The scheme optimizes the expected mean squared error of prediction in validation samples. (Ogawa 1987) proposed a non-overlapping hierarchical clusterwise regression procedure that allows for concurrent estimation and segmentation using logit estimation. (Kamakura 1988) proposed a similar methodology for conjoint models using least squares estimation. (Wedel and Kistemaker 1989) proposed a generalization of the clusterwise regression to handle more than one observation per individual and which yields nonoverlapping, nonhierarchical segments. (DeSarbo, Oliver and Rangaswamy 1989) proposed an overlapping nonhierarchical clusterwise regression method that uses a simulated annealing algorithm for optimization. (Wedel and Steenkamp 1989, 1991) proposed a fuzzy nonhierarchical clusterwise regression algorithm that permits subjects to have partial membership in at least one segment.

Probably, the advent of latent class and finite mixture models stands out to be the most far-reaching development in market segmentation. The merit of these models is that they allow for simultaneous segmentation, estimation and enable statistical inference. Work on latent class models was initiated by (Quandt 1972) who introduced the concept of switching regression models. (Goldfield and Quandt 1973, 1976) proposed a hidden Markov switching regression approach in which membership of observations within a cluster is modelled by a Markov process. (Engel and Hamilton 1990) extended the switching regression method to time series. The models describe discrete shifts in autoregressive parameters, where the shifts themselves are modelled by a hidden discrete-time Markov process. (DeSarbo, Wedel, Vriens, Ramaswamy 1992) and (Wedel and DeSarbo 1995) proposed a multivariate normal latent class model using the EM algorithm which calculates the posterior probabilities in the E-step. In an excellent review, (Vriens, Wedel and Wilms 1996) conducted a Monte Carlo comparison of several traditional and integrated conjoint segmentation methods. The authors found that Latent Class segmentation models performed best in terms of parameter recovery, segment membership recovery and predictive accuracy.

3. A LATENT CLASS MODEL

The latent class model described below was proposed by (DeSarbo, Wedel, Vriens, Ramaswamy 1992) in which market segments and part-worth utilities are estimated simultaneously using mixtures of multivariate conditional normal distributions. Parameters of these mixtures are estimated using the EM algorithm. Given that the n^{th} respondent belongs to the k^{th} segment, the conditional multivariate density of the dependent vectors $\mathbf{y}_n = \left(y_{nj}\right)$ for $j = 1,...,J$ replications is:

$$f_{n|k}(\mathbf{y}_n;\boldsymbol{\beta}_k) = (2\pi)^{-J/2}\left|\boldsymbol{\Sigma}_k\right|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{y}_n - \mathbf{X}\boldsymbol{\beta}_k)'\boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_n - \mathbf{X}\boldsymbol{\beta}_k)\right]$$

where $\boldsymbol{\Sigma}_k$ is the variance-covariance matrix of $\mathbf{y}_n$ given segment k. The unconditional density function is:

$$f_n\left(\mathbf{y}_n;\boldsymbol{\pi},\boldsymbol{\beta}\right) = \sum_1^K \pi_k \cdot f_{n|k}\left(\mathbf{y}_n;\boldsymbol{\beta}_k\right)$$

Using a likelihood approach, the log likelihood function can be formulated as follows:

$$\ln L\left(\boldsymbol{\pi},\boldsymbol{\beta}\right) = \ln\prod_{n=1}^{N} f_n\left(\mathbf{y}_n;\boldsymbol{\pi},\boldsymbol{\beta}\right) = \sum_{n=1}^{N}\ln\sum_{k=1}^{K}\pi_k \cdot f_{n|k}\left(\mathbf{y}_n\left|\boldsymbol{\beta}_k\right.\right)$$

Maximizing the expected log-likelihood function is not an easy task. An effective procedure that fits a latent class model with K segments is to maximize the expected complete log-likelihood function using the EM algorithm. The idea behind the EM algorithm is to augment the observed data by introducing unobserved 0-1 indicators λ_{nk}, indicating whether the n^{th} respondent belongs to the

k^{th} segment. Given the matrix $\mathbf{\Lambda} = (\lambda_{nk})$ of unobserved data, the complete log-likelihood function is:

$$\ln L(\boldsymbol{\pi}, \boldsymbol{\beta} | \mathbf{\Lambda}) = \sum_{n=1}^{N} \sum_{k=1}^{K} \lambda_{nk} . \ln f_{n|k}(\mathbf{y}_n | \boldsymbol{\beta}_k) + \sum_{n=1}^{N} \sum_{k=1}^{K} \lambda_{nk} . \ln(\pi_k)$$

$\ln L(\boldsymbol{\pi}, \boldsymbol{\beta} | \mathbf{\Lambda})$ has a simpler form than $\ln L(\boldsymbol{\pi}, \boldsymbol{\beta})$ and is easy to maximize. Once the parameter $\boldsymbol{\beta}_k$ and π_k are estimated, the posterior probability $\hat{p}_{nk} = E(\lambda_{nk})$ can be calculated using Bayes' theorem.

$$\hat{p}_{nk} = E(\lambda_{nk}) = \frac{\hat{\pi}_k \cdot f_{n|k}(\mathbf{y}_n | \hat{\boldsymbol{\beta}}_k)}{\sum_1^K \hat{\pi}_k \cdot f_{n|k}(\mathbf{y}_n | \hat{\boldsymbol{\beta}}_k)} \quad \text{where} \quad \sum_{i=1}^{K} \hat{p}_{nk} = 1$$

The iterative procedure is initiated by first setting pseudo random real values to $\hat{p}_{nk}$ in the range [0-1]. The EM algorithm updates alternately the parameters $\hat{\boldsymbol{\beta}}_k, \hat{\pi}_k$ and the posterior probabilities $\hat{p}_{nk}$ until it converges. Subjects are then assigned to the segment with highest posterior probability $\hat{p}_{nk}$.

The Bayesian information criterion (BIC) will be used in this latent class model to identify the number of segments.

$$BIC = -2 \log L + d \log N$$

d is the number of estimated parameters and N is the number of respondents.

4. APPLICATION

The main objective of this study is to establish which factors influence consumers' choices when buying mobile phones; which characteristics of the mobile phones are identified as most important by consumers in the market-place. What feature of the product effectively improves market sales? Do consumers give more priority to price or to brand? These are some of the questions that will be addressed in this paper. The four selected mobile phone attributes included brand (A and B), price (€150, €175 and €200), whether the mobile phone has internet access and touch screen facility. By choosing a complete factorial design, twenty-four profiles of different mobile phones were generated using a full profile approach. The stimuli were described using a verbal approach by providing details about each attributes. The questionnaire was sent to a 778 university students using an online survey. The respondents had to rate each profile using a 7-point Likert scale, where 1 corresponds to an unworthy mobile phone and 7 corresponds to a very worthy one. The participants were also asked to specify their gender, age and number of mobile phones owned.

The latent class model included all four item-attributes and three individual covariates. Since some of the predictors are categorical and others are continuous, a mixed model was assumed since it allows some attributes to follow the part-worth model while others follow the vector model. To identify the optimal number of segments, the latent class model was fitted several times each time changing the number of segments from 1 to 3. For each solution the BIC criterion was computed. Table 1 displays that the two-segment solution is the one which minimizes the criterion.

Number of segments K	Deviance (-2 log L)	Number of parameters d	BIC
1	43192	9	43252
2	41503	18	41623
3	41464	27	41644

Table1: BIC value for each segment solution

For each respondent, two posterior probabilities were computed which provided the probabilities that the respondent belonged to segment 1 and 2. The algorithm then allocated each respondent to the segment with highest posterior probability. In the two-segment model, 467 respondents were allocated to segment 1 and 311 subjects were allocated to segment 2.

5. RESULTS OF LATENT CLASS ANALYSIS

Table 2 displays the parameter estimates and standard errors for each segment solution.

Term	Segment 1		Segment 2	
	Par. Est.	St. Error	Par. Est.	St. Error
Constant	3.752	0.091	3.153	0.081
Brand (A)	1.307	0.023	0.069	0.078
Brand (B)	alias	alias	alias	alias
Price (€150)	0.118	0.034	1.942	0.118
Price (€175)	0.085	0.027	1.627	0.121
Price (€200)	alias	alias	alias	alias
Int. access (Yes)	1.452	0.026	0.140	0.103
Int. access (No)	alias	alias	alias	alias
Touch screen (Yes)	1.358	0.033	-0.123	0.104
Touch screen (No)	alias	alias	alias	alias
Gender (Male)	-0.014	0.042	0.013	0.079
Gender (Female)	alias	alias	alias	alias
Age	-0.007	0.052	-0.014	0.045
No. of mobiles	-0.115	0.079	-0.037	0.083

Table 2: Parameter estimates and standard errors

Respondents in segment 1 have strong brand preferences but do not consider the price as a monetary constrain. These respondents give more worth to brand A mobile phones having touch screen facility and internet access. On the other hand, respondents in segment 2 are price sensitive but hardly discriminate between the brands. These respondents do not value much any of the mobile phone facilities and see no bargain in buying expensive phones. In both segments, the worth of mobile phones tends to decrease with an increase of user's age and an increase in the number of mobile phones owned by user; however, both predictors are not significant at the 0.05 level of significance. The mean rating scores provided by males and females varied marginally across the levels of brand, price, internet access and touch screen facility.

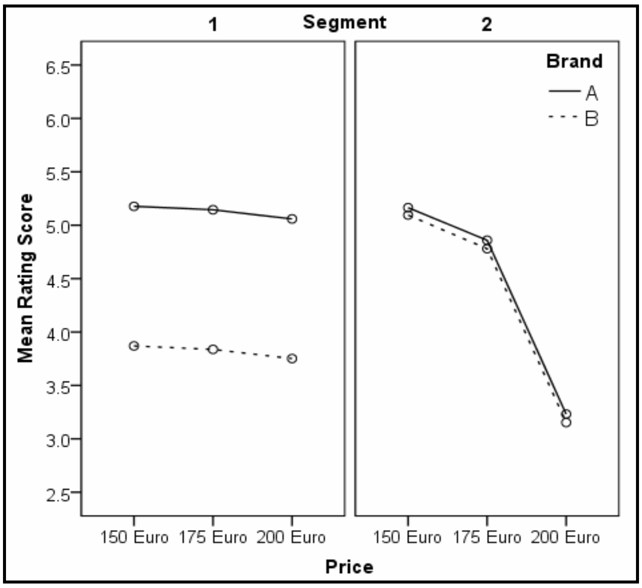

Figure 1: Mean worth by segment, price and brand

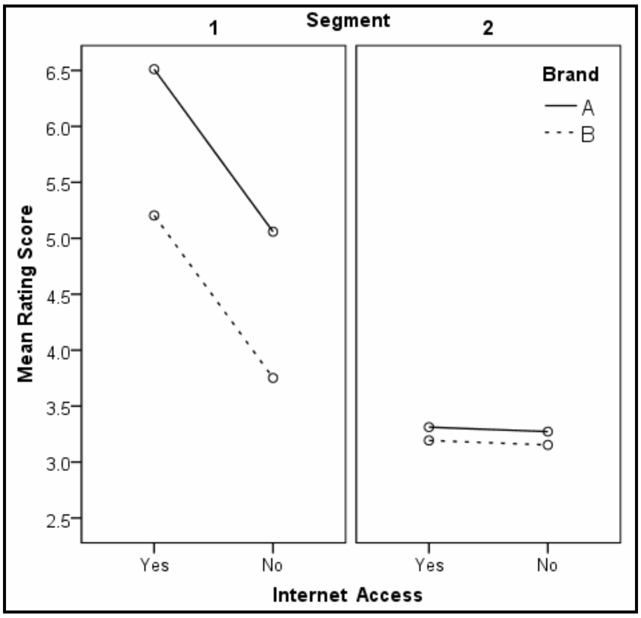

Figure 2: Mean worth by segment, internet access, brand

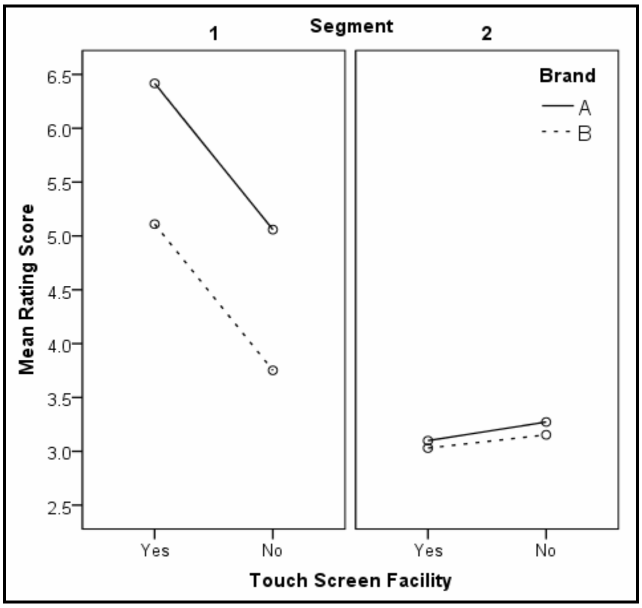

Figure 3: Mean worth by segment, screen facility, brand

REFERENCES

DeSarbo, W.S., Oliver, R.L. and Rangaswamy, A. (1989), A Simulated Annealing Methodology for Clusterwise Linear Regression, *Psychometrika*, 54, 707-736.

DeSarbo, W.S., Wedel, M., Vriens, M. and Ramaswamy, V. (1992), Latent Class Metric Conjoint Analysis, *Marketing Letters*, 3, 273-288.

Johnson, R.M. (1974), Trade-Off Analysis of Consumer Values, *Journal of Marketing Research*, 11, 121-127.

Green, P.E and Rao, V.R. (1971), Conjoint Measurement for Quantifying Judgmental Data, *Journal of Market Research*, 8, 355-363.

Hagerty, M.R. (1985), Improving the predictive Power of Conjoint Analysis: Use of Factor and Cluster Analysis *Journal of Marketing Research*, 22, 168-184.

Kamakura, W.A. (1988), A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments, *Journal of Marketing Research*, 25, 157-167.

Ogawa, K. (1987), Approach to Simultaneous Estimation and Segmentation in Conjoint Analysis, *Marketing Science*, 6, 66-81.

Srinivasan, V. and Wyner, G.A. (1989), CASEMAP: A computer assisted Self-Explication of multi attributed preference, *Product Development and Testing* 91-111.

Vriens, M., Wedel, M. and Wilms, T. (1996), Conjoint Segmentation Methods A Monte Carlo Comparison, *Journal of Marketing Research*, 23, 73-85.

Wedel, M. and DeSarbo, W.S. (1995), A Mixture Likelihood Approach for Generalized Linear Models, *Journal of Classification*, 12, 1-35.

Wedel, M. and Kistemaker, C. (1989), Consumer Benefit Segmentation using Clusterwise Linear Regression, *Journal of Research Marketing*, 6, 45-49

Wedel, M. and Steenkamp, J.B. (1989), Fuzzy Clusterwise Regression Approach to Benefit Segmentation, *Journal of Research in Marketing*, 6, 241-258.

Wedel, M. and Steenkamp, J.B. (1991), Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Segmentation, *Journal of Market Research*, 28, 385-396.

AUTHOR BIOGRAPHY

LIBERATO CAMILLERI studied Mathematics and Statistics at the University of Malta. He received his PhD degree in Applied Statistics in 2005 from Lancaster University. His research specialization areas are related to statistical models, which include Generalized Linear models, Latent Class models, Multilevel models and Mixture models. He is presently a senior lecturer in the Statistics department at the University of Malta.