SEGMENTING PREFERENCES FOR INVESTMENT BONDS USING LATENT VARIABLE MIXTURE MODELS

Liberato Camilleri

Department of Statistics and Operations Research; University of Malta

Malta

liberato.camilleri@um.edu.mt

Helena Francalanza*

Department of Mathematics; Junior College

Malta

helena.francalanza@um.edu.mt

Abstract

Market segmentation is a key component of conjoint analysis which addresses consumer preference heterogeneity. Members in a segment are assumed to be homogenous in their views and preferences when worthing an item but distinctly heterogenous to members of other segments. Latent class methodology is one of the several conjoint segmentation procedures that overcome the limitations of aggregate analysis and a-priori segmentation. The main benefit of Latent class models is that market segment membership and regression parameters of each derived segment are estimated simultaneously. The Latent class model presented in this paper uses mixtures of multivariate conditional normal distributions to analyze rating data, where the likelihood is maximized using the EM algorithm. The application focuses on customer preferences for investment bonds described by four attributes; currency, coupon rate, redemption term and price. A number of demographic variables are used to generate segments that are accessible and actionable.

Key words: Latent class models, EM algorithm, Market Segmentation, Conjoint Analysis

Topic Group: Marketing and Consumer behaviour

INTRODUCTION

Market segmentation has become a dominant concept in marketing practice. Besides being one of the major ways of operationalizing the marketing concept, segmentation provides guidelines for a firm's marketing strategy and resource allocation to increase the expected

profitability (Wind 1978). Understanding the diversity of preferences and sensitivities of customers in the market is one of the greatest challenges of market research. Market segmentation describes the division of a market into homogenous clusters, whose members respond differently to promotion, advertising, communication and other marketing variables. These clusters are created to group customers with similar needs, tastes and preferences, so that products or services can be optimally designed and targeted.

Market segmentation was first described by (Smith 1956) who recognized that segments are directly derived from the diversity of customer wants. The market environment is not static and market segments change composition over time. It is the interest of every market researcher to identify these changes. With more direct access to customers via databases, the market environment presents new challenges and opportunities for market segmentation. New developments in information technology provide marketers with much richer information on consumer behaviour. The rapid growth of new technologies in information, development of product, production and distribution enables a new company to make more efficient use of marketing resources, focussing on the best segments for its market products. The ability of a firm to differentiate its products relative to competing firms is essential for its survival. This survival depends on finding and addressing a niche rather than trying to be all things to all consumers. Consequently, marketers are focussing on smaller segments with micro marketing and direct marketing approaches. On the other hand, the increasing globalisation of most product markets is leading many multi-product manufacturers to look at global markets that cut across continents.

Six criteria have been frequently put forward as being essential for effective and profitable marketing strategies. The **identifiablity** criterion is the extent that marketers identify differences between distinct groups of customers in the market and the ability to classify each customer into one or more segments. The **substantiality** criterion refers to the size issue. If the identified segment is large enough to ensure profitability then it warrants separate market targeting. In micro markets and mass customisation, smaller segments become profitable due to lower marginal marketing costs; whereas in direct marketing, the criterion of substantiality can be applied to each individual customer. The **accessibility** criterion is the degree to which marketers are able to reach the targeted segment by a distinct marketing mix strategy. Once segments are identified and products are designed to suit their tastes, the marketer must be able to identify members of the segments so that marketing efforts can be directed to them. In other words, the message must reach the right market segment by using the right promotional strategies, media sources and distributional efforts to target these people. The **responsiveness** criterion is the degree by which segments respond uniquely to marketing effort targeted at them. Responsiveness is crucial for the effectiveness of any market segmentation strategy. Once the market is segmented, **stability** is necessary in which the segments do not change their composition or behaviour during the period for the identification of members and the implementation of the segmented market strategy. It is very likely that the segment will not be viable if its existence is the result of a short-term phenomenon,. The **actionability** criterion refers to the extent to which the identified market segments provide direction of marketing efforts. Segments are actionable if their identification provides guidance for decisions on the effective specification of associated marketing strategies towards segment targets.

CONJOINT SEGMENTATION METHODS

In most of the traditional a-priori segmentation approaches the type and number of segments were determined in advance by the researcher in which consumers were, very often, assigned to segments on the basis of demographic and socio-economic variables. Subsequently segmentation shifted to post-hoc predictive approaches because its recent developments allow for the grouping of consumers according to how they respond to product features in making choice decisions. Segmentation methods differ in three aspects: the type of partitioning assumed; the algorithms and estimation procedures used and the criterion being optimized. Some of these conjoint segmentation methods are summarized in the subsequent section.

A review of segmentation methods

A conjoint segmentation procedure proposed by (Green and DeSarbo 1979) is componential segmentation in which consumer descriptive variables are used. Consumer profiles are first generated on the basis of such characteristics. Respondents matching these profiles are chosen from a sample frame and asked to complete a conjoint task. From these evaluations, the componential segmentation model estimates both the main effects of the design variables and interactions between design variables of product and subject profiles. Estimation is carried out by minimizing the error sum of squares and the segmentation scheme is non-overlapping.

In the traditional two-stage conjoint segmentation approach estimation and clustering are conducted consecutively. Individual-level parameter estimates are first obtained by using least squares regression. At the second stage, subjects are clustered into segments on the basis of similarity of the estimated parameters through hierarchical or nonhierarchical nonoverlapping clustering procedures. One of the limitations of this two stage approach is that it ignores biasing errors. A second problem is that the use of fractional factorial designs often leaves few degrees of freedom for estimation at the individual level. This makes the parameter estimates unreliable as they become more sensitive to the measurement error. A third problem arises when the predictors are collinear. Near linear dependencies render it more difficult to sort out the impact of each predictor on the response and parameters estimates tend to be unreliable. This in turn may cause misclassification of individuals and negatively affects the goodness-of-fit and the power of the significance tests. A fourth limitation is that least squares regression and clustering procedures optimize different criteria.

(Green and Srinivasan 1978) proposed an alternative two-stage procedure. In the first step, consumers are clustered on the basis of their preference ratings whereas in the second step, separate conjoint models are estimated across the subjects in each of the identified segments. So rather than clustering consumers on the basis of similar parameter estimates at individual level, this method applies regression to the responses in each cluster to obtain more reliable parameter estimates. This procedure in effect increases the number of observations available for estimating the parameters and thus reduces the error of estimation.

(Hagerty 1985) proposed a method based on a weighting scheme which represents a factor-type partitioning of the sample. This weighting scheme presents an optimal overlapping partitioning obtained by a Q-factor analysis of the between-subject correlation matrix of preferences. A possible problem with this method is the interpretation of the factor solution

in terms of segments. The number of extracted factors need not be an adequate indicator of the number of segments. Another problem is that the factoring solutions are not unique, given the rotational indeterminacy of such factor solutions. Procedures proposed to identify segments on the basis of the factor solution, very often, result in a loss in predictive accuracy.

In response to the limitations of a-priori and two-stage procedures 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. (Kamakura 1988) suggested a hierarchical clusterwise regression procedure that allows for prediction within segments. At the first stage of the algorithm a regression equation is estimated for each subject using ordinary least squares, yielding regression parameter estimates of several independent variables for each subject. In the second stage a weighting scheme is devised that group subjects to maximise the accuracy with which preferences are predicted from product profiles. The fusion of any two subjects that yields the minimum increase in the total residual sum of squares of the regression across all clusters is retained and the two subjects are combined. The agglomerative process is similar to that of Ward's method. In each successive stage, segments that provide the smallest possible increase in the pooled within-segment error variance are linked together. A predictive accuracy index is computed at each aggregation level and provides an intuitive criterion for deciding how many segments to retain. There are two disadvantages to this agglomerative hierarchical method. First, the clustering process implies that this method depends in the initial stages on parameter estimates at the individual level, thereby creating the danger of misclassification at an early stage due to unreliable estimates. This misclassification may extend to higher levels in the hierarchical clustering process. Second, the number of parameters at the individual level may exceed the number of responses and so cannot be estimated. Models that are over-parameterised at the individual level yield unstable individual parameter estimates due to lack of degrees of freedom. Statistical tests to check for the significance of parameter estimates and check for homogeneity within the segments cannot be used because the asymptotic properties do not apply when the number of estimated parameters is close to the number of observations.

(Ogawa 1987) presented an approach for rank order preferences that employs simultaneous segmentation and estimation of conjoint models by using a hierarchical, non-overlapping clustering method. His formulation employs a stochastic logit framework. To evade problems with uniqueness of parameter estimates the author proposed a ridge regression-like procedure to estimate parameters at the individual level using multinomial logit models. An information criterion is also proposed to aggregate consumers hierarchically. This agglomerative method starts with single subject clusters and segments are combined iteratively to give a minimum reduction of the aggregate log-likelihood.

Several non-hierarchical procedures based on optimisation criteria are descriptive clustering methods that do not distinguish between dependent and independent variables. (Spath 1979, 1982) proposed a clusterwise linear regression procedure to find homogeneous groups in terms of the relationship between dependent and independent variables and simultaneously estimate corresponding regression functions within the clusters such that the sum of the error sums of squares over all clusters is minimized. Spath's method handles only one observation per individual. (Wedel and Kistemaker 1989) proposed a generalization of clusterwise regression by extending Spath's method to handle more than one observation per individual and which estimates parameters and segments simultaneously. Their

procedure uses (Banfield and Bassil 1977) exchange algorithm to maximize the likelihood and yields nonoverlapping, nonhierarchical segments.

(DeSarbo et al., 1989) proposed an overlapping clusterwise regression procedure that uses a simulated annealing algorithm for optimisation. This methodology can accommodate more general clusterwise linear regression formulations. It allows for multiple dependent variables, replicated observations by respondent, overlapping and nonoverlapping clusters and constraints on cluster membership. Computationally, simulated annealing has been devised as a general optimisation methodology to find the global optimum of a function that may have several local optima. This technique is based on a controlled random search that samples the objective function in a feasible region of the parameter space. The simulated annealing procedure starts from a random initial partition of the sample, and iteratively specifies steps in a random direction in the parameter space. If the new value of the objective function improves the criterion then the new solution is accepted. If the new value of the objective function does not improve the criterion then the new solution is rejected with a probability proportional to the decrease in the criterion value. The merit of this procedure is that it is less burdened with convergence to local optima.

(Wedel and Steenkamp 1989) proposed a fuzzy clusterwise regression algorithm that differs from the other fuzzy procedures since clusters are defined from regressions of the dependent variable on a set of explanatory variables. Similar to other fuzzy algorithms, partitioning of the data is carried out by minimizing the residual sum of squares criterion, which represents the sum of the distances of subjects from the regression equations in all clusters. The clustering algorithm iterates between two steps: computing regression parameters within each cluster and calculating fuzzy membership of subjects in clusters. (Wedel and Steenkamp 1991) generalizes this fuzzy clusterwise procedure to allow for a simultaneous grouping of both consumers and brands into groups, making possible the identification of market segments and market structures at the same time. There are two potential problems with this approach. The first is that the users must subjectively specify a fuzzy weight parameter that influences the degree of separation of the clusters and the second is that the statistical properties of the estimators are not established.

Probably, the advent of latent class models stands out to be the most extensive development in market segmentation. The works of (Wedel and DeSarbo 1995) and (DeSarbo et al., 1992) brought major changes in market segmentation applications. The major merit of these models is that they allow for simultaneous estimation and segmentation and enable correct statistical inference. 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.

THEORETICAL FRAMEWORK OF LATENT CLASS MODEL

One of the criteria for effective market segmentation is to identify differences between distinct groups of customers in the market and be able to classify each customer into a segment. The general principle of latent class models is that each segment defines a different probability structure for the response variable. For the segmentation procedure a latent class model with K segments is proposed.

$$H\left(\mathbf{y}_n ; \boldsymbol{\pi}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma}\right) = \sum_{k=1}^{K} \pi_k f_{nk}\left(\mathbf{y}_n \,\middle|\, \mathbf{X}, \boldsymbol{\beta}_k, \boldsymbol{\Sigma}_k\right)$$

$n = 1, \ldots, N$ respondents;

$k = 1, \ldots, K$ derived segments;

π_k is the proportion of respondents in segment k and $\boldsymbol{\pi} = \left(\pi_1, \ldots, \pi_K\right)$;

$\mathbf{y}_n$ is the vector of response ratings elicited by consumer n;

$\mathbf{X}$ is the data matrix;

$\boldsymbol{\beta}_k$ is the vector of parameter estimates for segment k and $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K\right)'$;

$\boldsymbol{\Sigma}_k$ is the covariance matrix estimated for segment k and $\boldsymbol{\Sigma} = \left(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K\right)'$.

It is assumed that $\sum_{1}^{K} \pi_k = 1$ and each f_{nk} has a conditional multivariate normal distribution.

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

The log-likelihood expression for N independent respondents is given by:

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

The derivatives of the expected log-likelihood function $E\left[\ln L\left(\boldsymbol{\pi}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma}\right)\right]$ with respect to the parameters are not straightforward. An effective procedure to fit a latent class model with K segments is to maximize the expected complete log-likelihood function using the iterative EM algorithm proposed by (Dempster et al., 1977). The idea behind the EM algorithm is to augment the observed data by introducing unobserved data λ_{nk}. This is a 0-1 indicator indicating whether respondent n is in segment k. Given the matrix $\boldsymbol{\Lambda} = \left(\lambda_{nk}\right)$ the complete log-likelihood function is given by:

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

$\ln L\left(\boldsymbol{\pi}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma} \,\middle|\, \boldsymbol{\Lambda}\right)$ has a simpler form than $\ln L\left(\boldsymbol{\pi}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma}\right)$ and the derivatives are manageable. Each iteration is composed of two steps - an E-step and an M-step. In the E-step, the expected log-likelihood function is calculated with respect to the conditional distribution of the unobserved data matrix $\boldsymbol{\Lambda} = \left(\lambda_{nk}\right)$ given the data and the provisional parameter estimates $\hat{\pi}_k$, $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\Sigma}}_k$. This is carried out by replacing $E\left(\lambda_{nk}\right)$ by the posterior probabilities $\hat{p}_{nk}$

$$E\left[\ln L\left(\boldsymbol{\pi},\mathbf{X},\boldsymbol{\beta},\boldsymbol{\Sigma}\mid\boldsymbol{\Lambda}\right)\right]=\sum_{n=1}^{N}\sum_{k=1}^{K}\hat{p}_{nk}.\ln f_{nk}\left(\mathbf{y}_{n}\mid\mathbf{X},\boldsymbol{\beta}_{k},\boldsymbol{\Sigma}_{k}\right)+\sum_{n=1}^{N}\sum_{k=1}^{K}\hat{p}_{nk}.\ln\left(\pi_{k}\right)$$

$$\hat{p}_{nk}=E\left(\lambda_{nk}\right)=\frac{\hat{\pi}_{k}.f_{nk}\left(\mathbf{y}_{n}\mid\mathbf{X},\hat{\boldsymbol{\beta}}_{k},\hat{\boldsymbol{\Sigma}}_{k}\right)}{\sum_{k=1}^{K}\hat{\pi}_{k}.f_{nk}\left(\mathbf{y}_{n}\mid\mathbf{X},\hat{\boldsymbol{\beta}}_{k},\hat{\boldsymbol{\Sigma}}_{k}\right)}\ \text{ and }\ \sum_{k=1}^{K}\hat{p}_{nk}=1$$

In the M-step the two terms of $E\left[\ln L\left(\boldsymbol{\pi},\mathbf{X},\boldsymbol{\beta},\boldsymbol{\Sigma}\mid\boldsymbol{\Lambda}\right)\right]$ are maximized separately with respect to the parameters π_{k} and $\boldsymbol{\beta}_{k}$. Maximizing the first term of the expected log-likelihood function with respect to $\boldsymbol{\beta}_{k}$ leads to independently solving each of the K expressions

$$\sum_{n=1}^{N}\hat{p}_{nk}.\frac{\partial}{\partial\boldsymbol{\beta}_{k}}\ln f_{nk}\left(\mathbf{y}_{n}\mid\mathbf{X},\boldsymbol{\beta}_{k},\boldsymbol{\Sigma}_{k}\right)\ \text{ for }\ k=1,2,...,K$$

Maximizing the second term of the expected log-likelihood function with respect to π_{k}, subject to the constraint $\sum_{k=1}^{K}\pi_{k}=1$, yields

$$\hat{\pi}_{k}=\frac{1}{N}\sum_{n=1}^{N}\hat{p}_{nk}\ \text{ for }\ k=1,2,...,K$$

METHODOLOGY AND IMPLEMENTATION

The EM algorithm for fitting latent class models is implemented as a set of GLIM macros. This is equivalent to the iterative fitting of a weighted generalized linear model with posterior probabilities recalculated at each iteration. The iterative procedure is initiated by setting random values to $\hat{p}_{nk}$. The algorithm then alternately updates the parameters $\hat{\pi}_{k}$, $\hat{\boldsymbol{\beta}}_{k}$ and $\hat{\boldsymbol{\Sigma}}_{k}$ and the probabilities $\hat{p}_{nk}$ until the process converges. The assignment of individuals to segments is done probabilistically by Bayes' Theorem. Individuals are assigned to the segment with the highest posterior probability $\hat{p}_{nk}$.

A problem associated with the application of the EM algorithm to latent class models is its convergence to local maxima. It is caused by the likelihood being multimodal, so that the algorithm becomes sensitive to the starting values used. A procedure that widens the search for the global maximum is to perturb the posterior probabilities at each iteration by adding to each probability a pseudo-random real value multiplied by a scalar. The pseudo-random real values are generated from a uniform distribution in the range [0,1] from and the scalar is initially set to 0.1. These modified posterior probabilities are rescaled such that they sum to 1 across the segments. This scalar is reduced systematically after a number of iterations so that the iterative procedure will eventually converge.

APPLICATION

Bond investment strategies vary between policyholders. Some policyholders may give more relevance to coupon rate and duration of investment; whereas other policyholders may give more importance to the issue price and credit rating of the issuer. Indeed the choices policyholders make depend upon several factors, which include their knowledge, their time frames, investment goals and the amount of risk that they are willing to take.

To illustrate the methodology a conjoint study on 300 policyholders was conducted to investigate their preference for investment bonds. Four product attributes, which included issue price, duration of investment, coupon rate and credit rating of issuing company were identified as being key determinant attributes. The study compared three issue prices (97, 100 and 103) with two investment periods (10 and 20 years) with three coupon rates (3%, 4% and 5%) and with two levels of credit rating (A and B). A complete factorial design was utilized which included 36 combinations of attribute manifestations. For data collection a full profile approach was used in which all the profiles had a unique attribute combination for an investment bond. Such a design guarantees orthogonality (independence) of the attributes and eventually this will result in an efficient estimation of the parameters. To reduce information overload on respondents two blocks of cards were presented and each respondent was handed a set of 18 cards with random assignment to block. Preference ratings were measured on a seven point scale where 1 corresponds to 'worst' and 7 corresponds to 'best'. A rating scale was chosen over a ranking scale on the merit that it express more the intensity of a preference. The data collection procedure used was person-to-person interview as this ensured a higher return rate. The linear predictor which relates the expected worth of an investment bond to its product attributes includes all main effects and all pairwise interactions. To make the derived market segments more accessible and actionable two subject profiles were also recorded including gender and their knowledge about investment bonds. The sample of 300 participants comprised equal number of males and females and equal number of participants with good knowledge and limited knowledge about investment bonds. Most respondents with good knowledge were employees in the financial sector.

Latent class models assume that observed data is made up of several unknown homogeneous segments which are mixed up in an unknown proportion. The first statistical objective is to discover the true number of segments. To address this issue, three criteria were used to identify the correct number of homogeneous groups of respondents in a heterogeneous population. Two of these information criteria are based on the bias-corrected log-likelihood.

$$C = -2\log L(\Psi) + dc$$

d is the number of estimated parameters and c is a penalty constant and measures the complexity of the model. For the Akaike information criterion (AIC), $c = 2$ and for the Bayesian information criterion (BIC), $c = \ln(N)$, where N is the sample size. The third criterion includes an additional entropy term which is related to posterior probabilities $\hat{p}_{nk}$.

$$EN(\hat{p}_{nk}) = -\sum_{k=1}^{K}\sum_{n=1}^{N} \hat{p}_{nk} \log(\hat{p}_{nk})$$

This criterion, which is an approximation of the Integrated Classification Likelihood (ICL), assesses the degree of separation between the segments and is more appropriate for large cluster sizes and attempts to overcome the short-comings of AIC and BIC.

$$ICL = -2\log L(\Psi) + 2EN(\hat{p}_{nk}) + d\log(N)$$

This latent class model was fitted three times varying the number of segments from one to three clusters. To overcome the problem of convergence to local optima, ten different random starting values were considered for each model fit. The solution with the smallest log-likelihood was selected. The entropy and the number of estimated parameters were also recorded for each solution to determine the optimal number of segments

FINDINGS

Table 1 shows that BIC and ICL reach a two-segment solution whereas AIC reaches a three-segment solution.   Many authors have observed that AIC tend to overestimate the correct number of segments. Since AIC does not penalize complex models as heavily as the other two criteria we opt for a two-segment solution. After assigning each respondent to the segment with highest posterior probability, these were then categorized by their gender and knowledge about investment bonds.

Table 1: Determination of the number of segments using AIC, BIC and ICL

Number of Segments	$-2\log L(\Psi)$	Number of parameters	Entropy	AIC	BIC	ICL
1	2342.2	19	106.92	2361.2	2450.6	2664.4
2	1262.7	38	42.29	1338.7	1479.4	1564.0
3	1184.9	57	40.76	1298.9	1510.0	1591.5

Table 2 shows a higher proportion of male and female participants in segment 1 that have good knowledge about investment bonds; whereas segment 2 comprises a higher percentage of respondents with limited knowledge. The segments do not discriminate much between the gender groups.

Table 2: Number of respondents assigned to segments by gender and knowledge

Segment	Knowledge about investment bonds	Gender		Total
		Male	Female	
1	Good	48	46	94
	Limited	17	15	32
	Total	65	61	126
2	Good	25	31	56
	Limited	60	58	118
	Total	85	89	174

Figure 1 demonstrates that respondents in segment 1 discriminate between investment bonds having different coupon rates and different issue prices. Respondents in segment 1 worth bonds with low issue prices more than bonds with high issue prices. These respondents worth high-coupon rated bonds more than low-rated coupon bonds. On the contrary, respondents in segment 2 discriminate between different coupon rates but hardly differentiate between different issue prices.

Figure 1: Predicted rating scores by cluster allocation, coupon rate and issue price
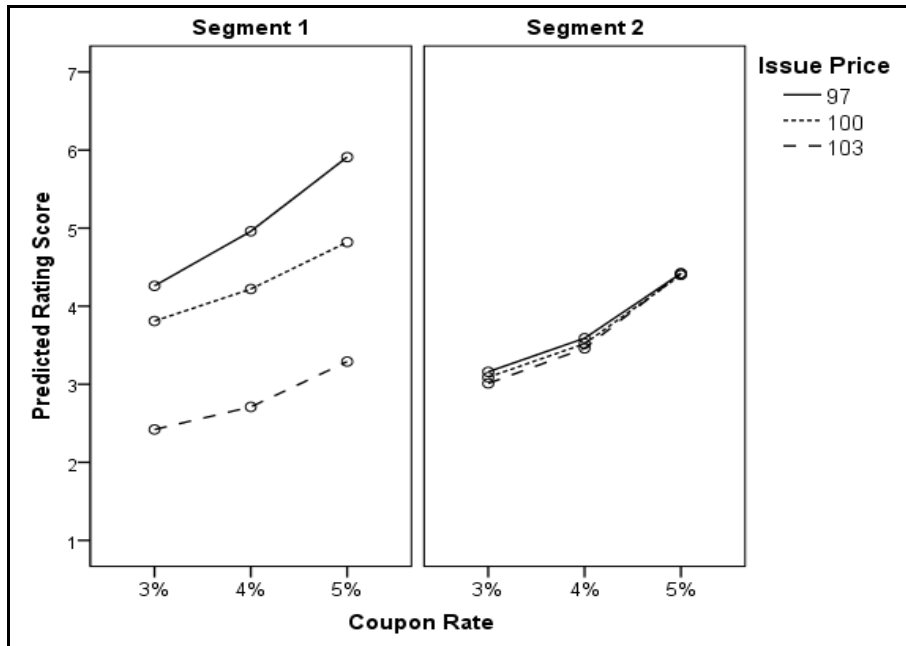

Figure 2 demonstrates that respondents in both segments worth bonds with a high coupon rate more than bonds with a lower coupon rate. Respondents in both segments worth bonds with an 'A' credit rating more than bonds with a 'B' credit rating; however, the difference in the expected worth for these two types of investment bonds is more conspicuous for respondents in segment 1. Respondents in segment 2 give more priority to the coupon rate than the credit rating of the issuing company.

Figure 2: Predicted rating scores by cluster allocation, coupon rate and credit rating
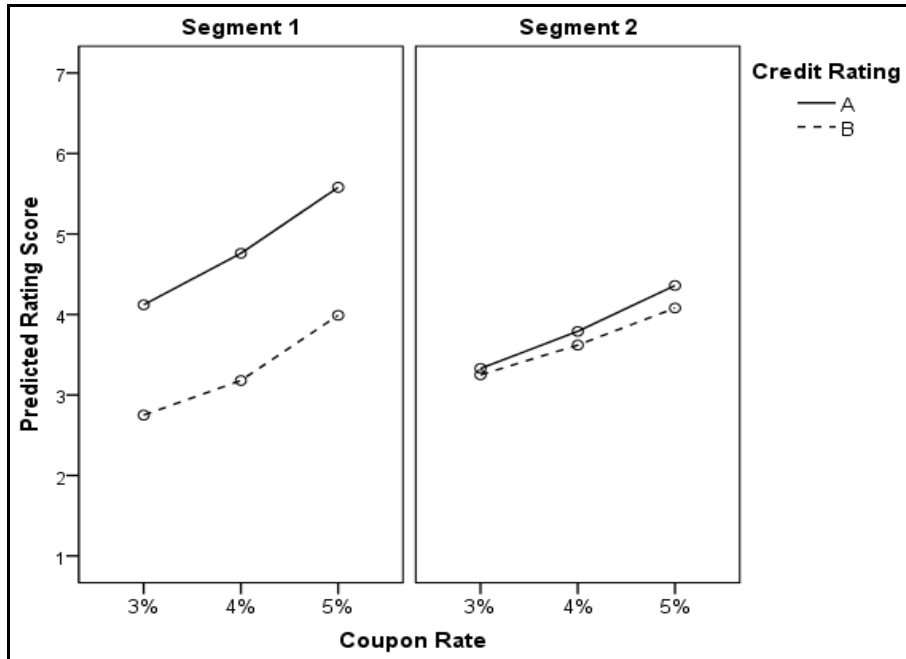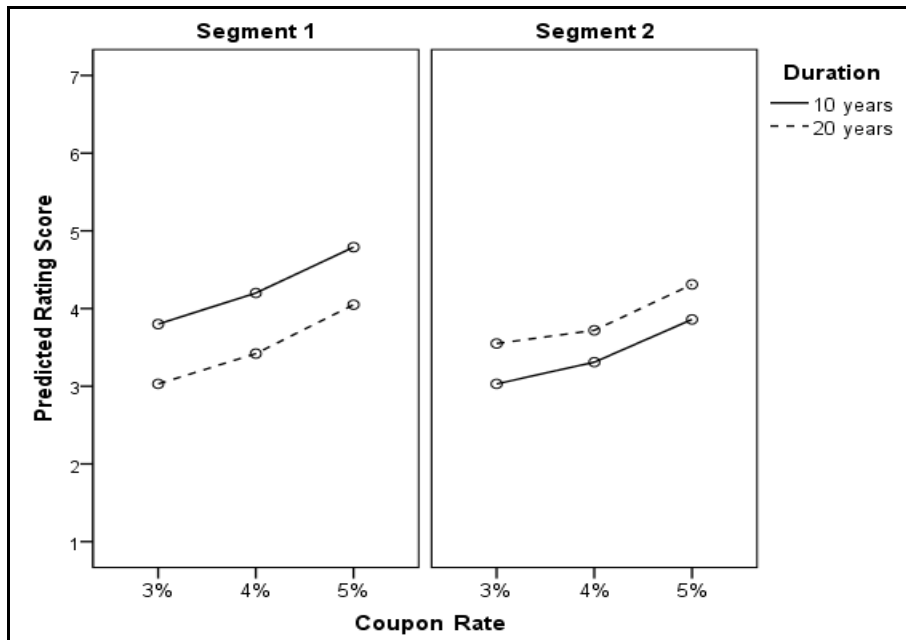

Figure 3 shows those respondents in segment 1 worth investments bonds issued for a short term more than long-term bonds. Conversely, respondents in segment 2 prefer to invest their money in bonds that are issued for longer durations.

Figure 3: Predicted rating scores by cluster allocation, coupon rate and duration

The main finding of the application related to investment bonds is that there are two groups of investors. Investors who have good knowledge of financial investments tend to give priority to all product attributes; whereas, investors with limited knowledge tend to give

more priority to coupon rates and dividend returns than to the issue price of the bond and credit rating of the issueing company. Moreover, investors with good knowledge of financial investments tend to invest their money for shorter terms than their counterparts with limited knowledge.

IMPLICATIONS FOR THEORY AND PRACTICE

Segmentation has proved to be a very useful concept to managers and modelling consumer heterogeneity is the central focus of many statistical marketing applications. Models that approximate market heterogeneity by a number of unobserved segments have great managerial appeal in many applications. Moreover, managers seem comfortable with the idea of market segments based on the assumption that consumers can be grouped into relatively homogeneous segments and the models appear to do a good job of identifying useful groups. Although segment-level models are very compelling from a managerial standpoint, they can over-simplify the market scenario and may have limited predictive validity. One of the major concerns underlined by market researchers is whether segment-level models enable marketers to customize their products or services to very small segments, particularly micro-marketing, direct marketing and mass customisation. Segment-level models may not be sufficiently accurate in estimating responses to marketing variables at the consumer level. The rapid growth of new technologies is enabling marketers to customize their products or services to very small segments where each consumer may represent a segment and the responses to marketing variables are estimated at the individual level. In other words, a set of idiosyncratic parameters is estimated for each subject, where the posterior distribution of individual-level parameters can be estimated using Bayesian methods. Bayesian estimation methods have gained popularity recently and their main advantage lies in obtaining posterior distributions of individual-level parameters based on the parameters of the prior distribution.

The majority of market research applications assume a discrete distribution. The popularity of this approach is partly due to the fact that the marginal likelihood is easily evaluated as a sum over a discrete number of mass points. (Heckman and Singer 1984) emphasize that any distribution can be approximated, to a high degree of accuracy, by a discrete distribution for a sufficient number of mass points. However, other authors argue that consumer heterogeneity is better described by a continuous rather than a discrete distribution. (Allenby and Rossi 1999), pointed out that the underlying assumption of a limited number of segments of individuals that are perfectly homogeneous within segments in finite mixture models is overly restrictive.nThe authors argue that by segmenting the market into a small number of homogeneous clusters leads to an artificial partition of the continuous distribution because a limited number of mass points may inadequately capture the full extent of heterogeneity in the data. (Lenk, DeSarbo, Green and Young 1996) argue that a discrete latent class or mixture approach to heterogeneity ignores the inherent differences across consumers and may result in a loss of predictive performance. In micro or direct marketing applications a continuous approximation of consumer heterogeneity may be more appropriate because targeting individual customers is more essential than identifying segments. (Allenby and Lenk 1994; Allenby and Ginter 1995) suggest that consumer preferences, tastes and response to marketing variables are distributed over the consumer population according to a continuous distribution rather than assuming a discrete distribution across homogeneous segments. Both discrete and continuous representations have advantages and disadvantages and there is no evidence that one representation outperforms the other. It is still an empirical question under which conditions one representation is more

appropriate than the other. (Arora, Allenby and Ginter 1998; Lenk and DeSarbo 2000) have developed segmentation models that reach a compromise between the two philosophies to account for both discrete segments and within segment heterogeneity. Their methodology combines discrete and continuous heterogeneity approaches. Both discrete and continuouos heterogeneity models can in principle be estimated with either maximum likelihood or Bayesian methods.

LIMITATIONS AND FUTURE RESEARCH

Our discussion focussed mainly on latent class models that address heterogeneity through a discrete distribution. These segment-level models assume that subjects within each segment respond similarly to a marketing mix but respond differently to others in other clusters. The main advantage of these models over traditional clustering techniques lies in simultaneous estimation and segmentation. The algorithm was devised to implement latent class models for normally-distributed rating responses, where the parameters are estimated using a maximum likelihood approach. The limitation of the algorithm is that the normality assumption may not always be adequate and inappropriate statistical assumptions may lead to deficiencies in the performance of the analyses, particularly when the distribution of rating scores is skewed. An alternative approach is to modify the algorithm by combining the Proportion odds model with the EM algorithm. The Proportion Odds model is more appropriate since it accommodates skewed ordinal categorical responses better when the normality assumption is not satisfied. Moreover, when the data set comprises response categories that have a natural ordering it is more sensible to work with cumulative link models since they utilize the ordering better.

Consider an ordinal scale from 1 to R and let Υ_j represent the R-category responses then $P(\Upsilon_j \leq r)$ can be defined as the rth cumulative probability of the jth item for $r = 1, 2, ..., R$. The cumulative probabilities reflect the ordering since;

$$P(\Upsilon_j \leq 1) \leq P(\Upsilon_j \leq 2) \leq ... \leq P(\Upsilon_j \leq R) = 1$$

Assuming that $\mathbf{X}$ is the design matrix containing the values of the explanatory variables, the jth row $\mathbf{x}_j$ of $\mathbf{X}$ contains the values of the explanatory variables for the jth item. Also, letting $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ be a vector of parameters for the explanatory variables and $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{R-1})$ be a vector of threshold parameters such that $\alpha_1 \leq \alpha_2 \leq ... \leq \alpha_{R-1}$, $\alpha_0 = -\infty$ and $\alpha_R = \infty$. The proportional odds model, which is the appropriate model for analyzing ordinal categorical responses is given by:

$$P(\Upsilon_j \leq r) = F(\alpha_r + \eta_i) \quad \text{for } r = 1, 2, ..., R-1$$

where $\eta_j = \mathbf{x}_j' \boldsymbol{\beta}$ and F is a cumulative distribution function. The link function F^{-1} is a strictly monotonic function in the range $[0,1]$ onto the real line. The cumulative link model links the cumulative probabilities $P(\Upsilon_j \leq r)$ to the real line, using the link function F^{-1}.

$$F^{-1}[P(\Upsilon_j \leq r)] = \alpha_r + \eta_j$$

$F(.)$ can be the logistic, normal or the extreme value distributions leading to logit, probit and complementary log-log link functions. The model suggested by (McCullagh 1980), for predicting the probabilities $\mu_j = P(\Upsilon_j = r)$ is given by

$$P(\Upsilon_j = r) = P(\Upsilon_j \le r) - P(\Upsilon_j \le r-1) = F(\alpha_r + \eta_j) - F(\alpha_{r-1} + \eta_j)$$

For the segmentation procedure a latent class model with K segments is considered

$$P(\Upsilon_j = r \mid \alpha, \beta, \pi) = \sum_{k=1}^{K} \pi_k P(\Upsilon_j = r \mid \alpha, \beta)$$

where $P(\Upsilon_j = r \mid \alpha, \beta) = F(\alpha_r + \mathbf{x}_j'\boldsymbol{\beta}) - F(\alpha_{r-1} + \mathbf{x}_j'\boldsymbol{\beta})$ and π_k is the proportion of respondents in the kth segment. The likelihood function is maximized using the EM algorithm which is equivalent to iterative fitting of a weighted GLM with posterior probabilities recalculated at each iteration. Moreover, the parameters within the segments are estimated at the same time that segment membership is identified.

REFERENCES

Allenby, G. M. & Ginter, J. L. (1995). Using Extremes to Design Products and Segment Markets, *Journal of Marketing Research*, 32, 392-405.

Allenby, G. M. & Lenk, P. J. (1994). Modelling Household Purchase Behaviour with Logistic Normal Regression, *Journal of the American Statistical Association* 89, 1218-1231.

Allenby, G. M. & Rossi, P. E. (1999). Marketing Models of Heterogeneity, *Journal of Econometrics*, 89, 57-78.

Arora, N., Allenby, G. M. & Ginter, J. L. (1998). A hierarchical Bayes model of primary and secondary demand. *Marketing Science*, 17, 29-44.

Bagozzi, R. P. (1994), *Advanced Methods of Marketing Research*, Blackwell Publisher.

Banfield, C. F. & Bassil, L. C. (1977). Transfer Algorithm for Nonhierarchical Classification, *Applied Statistics*, 26, 206-210.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM algorithm, *Journal of the Royal Statistical Society*, B, 39, 1-38.

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

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

Green, P. E. & DeSarbo, W. S. (1979). Componential Segmentation in the Analysis of Consumer Trade-offs; *Journal of Marketing*, 43, 83-91.

Green, P. E. & Srinivasan, V. (1978). Conjoint Analysis in Consumer Research: Issues and Outlook, *Journal of Consumer Research*, 5, 103-123.

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

Heckman, J. J. & Singer, B. (1984). A Method for Minimizing the Impact of Distributional Assumption in Econometric Models for Duration Data, *Econometrica*, 52, 271-320.

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

Lenk, P. J. & DeSarbo, W. S. (2000). Bayesian Inference for Finite Mixture of Generalized Linear Models with Random Effects, *Psychometrika*, 65, 93-119.

Lenk, P. J., DeSarbo, W. S., Green, P. E. & Young, M. R. (1996). Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs, *Marketing Science*, 15, 152-173.

McCullagh, P. (1980). Regression models for ordinal data. J. R. *Satist. Soc.* B 42, 109-142.

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

Rossi, P. E., Allenby, G. M. & McCulloch, R. (2006). *Bayesian Statistics and Marketing*, John Wiley and Sons Ltd.

Smith, W. (1956). Product Differentiation and Market Segmentation as Alternative Marketing Strategies, *Journal of Marketing*, 21, 3-8.

Spath, H. (1979). Clusterwise Linear Regression, *Computing*, 22, 367-373.

Spath, H. (1982). An Algorithm for Clusterwise Linear Regression, *Computing*, 29, 175-181.

Vriens, M., Oppewal, H. & Wedel, M. (1998). Ratings-based versus Choice-based Latent Class Conjoint Models – An empirical comparison, *Journal of the Market Research Society*, 40, 237-248.

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

Wedel, M. & Kamakura, W. A. (2000). *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers.

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

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

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

Wind, Y. (1978). Issues and Advances in Segmentation Research, *Journal of Market Research*, 15, 317-337.