Department of Mathematics
Faculty of Mechanical Engineering
Slovak University of Technology in Bratislava

6th International Conference
APLIMAT 2007

Segmenting the heterogeneity of tourist preferences using a latent class model combined with the EM algorithm

Camilleri Liberato, (M) and Portelli Mary Rose, (M)

Abstract. An important component of conjoint analysis is market segmentation where the main objective is to address the heterogeneity of consumer preferences. 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 they simultaneously estimate market segment membership and parameter estimates for each derived market segment. In this paper we present two latent class models. The first model is a latent class metric model using mixtures of multivariate conditional normal distributions to analyze rating data. The second is a latent class multinomial logit model used to analyze choice data. The EM algorithm is employed to maximize the likelihood in both models. The application focuses on tourists' preference and choice behaviour when assessing package tours. A number of demographic and product related explanatory variables are used to generate segments that are accessible and actionable. A Monte Carlo study is also presented in this paper. This study examines how the number of hypothetical subjects, number of specified segments and number of predictors affect the performance of the latent class metric conjoint model with respect to parameter recovery and segment membership recovery.

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

Mathematics Subject Classification: Primary 62P20; Secondary 91B26

1 Introduction

Conjoint analysis is a term used to define a wide range of techniques for estimating the worth customers give to the features or attributes that define a product or service. Today, conjoint analysis is an established technique for investigating customer preferences and has a tremendous effect on many aspects of business research. Since its introduction into marketing literature by (Green and Rao 1971) conjoint analysis has developed into a methodology for measuring customer preferences and predicting choice behaviour. The popularity of conjoint analysis hinges on the belief that it produces valid measurements of preferences as customers trade off between competing products. The application to marketing and business research has increased substantially over the last thirty-five years. This is reflected in the large variety of professionals - marketing research consultants, academics, software developers and industry professionals - who contributed to the development of conjoint analysis.

One of the objectives of marketers is to identify new products or modify existing ones. Another objective is to find the best target market for these products that will optimize profits. In an attempt to satisfy customer needs marketers are also interested in determining what products to offer, what prices to charge, how to promote the product and how to design and deliver the products to the customer. (Wittink and Cattin 1989) identified seven application fields in conjoint analysis. One of the most popular fields, which address customer preference heterogeneity, is market segmentation. In market segmentation a heterogeneous population of customers is represented as a collection of homogeneous subgroups where the customers in each cluster have similar needs and similar views of how to worth a product.

Traditionally segmentation procedures were carried out using either a-priori or two-stage methods. In a-priori segmentation approach the type and number of segments are determined in advance by the researcher. In a two-stage approach estimation and clustering are conducted 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 clustering algorithm or nonhierarchical (K-means) clustering procedures. 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. (Hagerty 1985) proposed a method based on a weighting scheme which represents a factor-type partitioning of the sample. This weighting scheme optimizes the expected mean squared error of prediction in validation samples. (Kamakura 1988) and (Ogawa 1987) proposed nonoverlapping hierarchical clusterwise regression procedures that allow for simultaneous segmentation and estimation of conjoint models. Kamakura uses least squares estimation and Ogawa uses logit estimation. (DeSarbo, Oliver and Rangaswamy 1989) proposed an overlapping nonhierarchical clusterwise regression procedure that uses a simulated annealing algorithm for optimization. (Wedel and Kistemaker 1989) proposed a generalization of clusterwise regression by extending (Spath 1979) method to handle more than one observation per individual and which yields nonoverlapping, nonhierarchical segments. Their procedure uses (Banfield and Bassil 1977) exchange algorithm to maximize the likelihood. (Wedel and Steenkamp 1989, 1991) proposed a fuzzy nonhierarchical clusterwise regression algorithm that permits customers to possess partial membership in several segments.

Probably, the advent of latent class and finite mixture models stands out to be the most far-reaching development in market segmentation. The works of (Wedel and DeSarbo 1995) and (DeSarbo, Wedel, Vriens and Ramaswamy 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 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.

2 A Latent Class Metric conjoint 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.

$$H\left(\mathbf{y}_n;\boldsymbol{\pi},\mathbf{X},\boldsymbol{\beta},\boldsymbol{\Sigma}\right)=\sum_{k=1}^{K}\pi_k f_{nk}\left(\mathbf{y}_n\left|\mathbf{X},\boldsymbol{\beta}_k,\boldsymbol{\Sigma}_k\right.\right) \tag{2.1}$$

$n=1,\ldots,N$ respondents;
$k=1,\ldots,K$ derived segments;
π_k is the proportion of respondents in segment *k* and $\boldsymbol{\pi}=\left(\pi_1,...,\pi_K\right)$;
$\mathbf{y}_n$ 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 $\boldsymbol{\beta}=\left(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_K\right)'$;
$\boldsymbol{\Sigma}_k$ is the covariance matrix estimated for segment *k* and $\boldsymbol{\Sigma}=\left(\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_K\right)'$.

It is assumed that $\sum_1^K \pi_k=1$ and each f_{nk} has a conditional multivariate normal distribution.

$$f_{nk}\left(\mathbf{y}_n\left|\mathbf{X},\boldsymbol{\beta}_k,\boldsymbol{\Sigma}_k\right.\right)=\left(2\pi\right)^{-J/2}\left|\boldsymbol{\Sigma}_k\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathbf{y}_n-\mathbf{X}\boldsymbol{\beta}_k\right)'\boldsymbol{\Sigma}_k^{-1}\left(\mathbf{y}_n-\mathbf{X}\boldsymbol{\beta}_k\right)\right] \tag{2.2}$$

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

$$\ln L\left(\boldsymbol{\pi},\mathbf{X},\boldsymbol{\beta},\boldsymbol{\Sigma}\right)=\ln\prod_{n=1}^{N}H\left(\mathbf{y}_n;\boldsymbol{\pi},\mathbf{X},\boldsymbol{\beta},\boldsymbol{\Sigma}\right)=\sum_{n=1}^{N}\ln\sum_{k=1}^{K}\pi_k.f_{nk}\left(\mathbf{y}_n\left|\mathbf{X},\boldsymbol{\beta}_k,\boldsymbol{\Sigma}_k\right.\right) \tag{2.3}$$

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. 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:

$$\ln L\left(\boldsymbol{\pi},\mathbf{X},\boldsymbol{\beta},\boldsymbol{\Sigma}\left|\boldsymbol{\Lambda}\right.\right)=\sum_{n=1}^{N}\sum_{k=1}^{K}\lambda_{nk}.\ln f_{nk}\left(\mathbf{y}_n\left|\mathbf{X},\boldsymbol{\beta}_k,\boldsymbol{\Sigma}_k\right.\right)+\sum_{n=1}^{N}\sum_{k=1}^{K}\lambda_{nk}.\ln\left(\pi_k\right) \tag{2.4}$$

$\ln L\left(\boldsymbol{\pi},\mathbf{X},\boldsymbol{\beta},\boldsymbol{\Sigma}\left|\boldsymbol{\Lambda}\right.\right)$ 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 $\mathbf{\Lambda} = (\lambda_{nk})$ 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 $E(\lambda_{nk})$ by the posterior probabilities $\hat{p}_{nk}$

$$E\left[\ln L\left(\boldsymbol{\pi}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma} \mid \mathbf{\Lambda}\right)\right] = \sum_{n=1}^{N}\sum_{k=1}^{K} \hat{p}_{nk}.\ln f_{nk}\left(\mathbf{y}_n \left|\mathbf{X}, \boldsymbol{\beta}_k, \boldsymbol{\Sigma}_k\right.\right) + \sum_{n=1}^{N}\sum_{k=1}^{K} \hat{p}_{nk}.\ln\left(\pi_k\right) \tag{2.5}$$

$$\hat{p}_{nk} = E\left(\lambda_{nk}\right) = \frac{\hat{\pi}_k . f_{nk}\left(\mathbf{y}_n \left|\mathbf{X}, \hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\Sigma}}_k\right.\right)}{\sum_{k=1}^{K} \hat{\pi}_k . f_{nk}\left(\mathbf{y}_n \left|\mathbf{X}, \hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\Sigma}}_k\right.\right)} \quad \text{and} \quad \sum_{i=1}^{K} \hat{p}_{nk} = 1 \tag{2.6}$$

In the M-step the two terms of $E\left[\ln L\left(\boldsymbol{\pi}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma} \mid \mathbf{\Lambda}\right)\right]$ are maximized separately with respect to the parameters π_k and $\boldsymbol{\beta}_k$. Maximizing the first term of Eq. (2.5) with respect to $\boldsymbol{\beta}_k$ leads to independently solving each of the K expressions

$$\sum_{n=1}^{N} \hat{p}_{nk} . \frac{\partial}{\partial \boldsymbol{\beta}_k} \ln f_{nk}\left(\mathbf{y}_n \left|\mathbf{X}, \boldsymbol{\beta}_k, \boldsymbol{\Sigma}_k\right.\right) \quad \text{for} \quad k = 1, 2, \ldots, K \tag{2.7}$$

Maximizing the second term of Eq. (2.5) with respect to π_k, subject to the constraint $\sum_{k=1}^{K} \pi_k = 1$, yields

$$\hat{\pi}_k = \frac{1}{N}\sum_{n=1}^{N} \hat{p}_{nk} \quad \text{for} \quad k = 1, 2, \ldots, K \tag{2.8}$$

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 Eq. (2.6). Individuals are assigned to the segment with the highest posterior probability $\hat{p}_{nk}$.

3. An Illustrative Application for Rating Data

The increasing number of tourist travelling to various destinations and the issues connected with different travel preferences makes the tourist a target of interest, especially for the travel agencies. To illustrate the methodology a conjoint study on 200 travellers was conducted to investigate tourist preferences for different package tours. Three key attributes that determine tourist choice behaviour are the cost of the package tour, the type of lodging offered and the city visited. The study compared four city destinations, four different prices and whether the accommodation was a single or a twin room. A complete factorial design was utilized which included 32 combinations of attribute manifestations. For data collection a full profile approach was used in which all the profiles (cards) had a unique attribute combination for a package tour. To reduce information overload on respondents two blocks of cards were presented and each respondent was handed a set of 16 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'. To make the derived market segments more accessible and actionable three demographic variables were recorded which included the age, gender and status of the respondents.

The utility function which relates the utility of a package tour to its attributes includes all main effects, a quadratic function of price and relevant pairwise interactions between item attributes and demographic variables. The quadratic function for price allows a dual role – tourists use the price both as a signal of the package tour quality and as a monetary constraint in choosing it. The EM algorithm for fitting latent class models was implemented as a set of GLIM macros.

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.

$$C = -2\log L(\mathbf{\Psi}) + dc \tag{3.1}$$

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, which uses the entropy

$$EN(\hat{p}_{nk}) = -\sum_{k=1}^{K}\sum_{n=1}^{N} \hat{p}_{nk} \log(\hat{p}_{nk}) \tag{3.2}$$

to assess the degree of separation between the segments, is an approximation of the Integrated Classification Likelihood (ICL).

$$ICL = -2\log L(\mathbf{\Psi}) + 2EN(\hat{p}_{nk}) + d\log(N) \tag{3.3}$$

This criterion is more appropriate for large cluster sizes and attempts to overcome the short-comings of AIC and BIC.

This latent class model was fitted four times varying the number of segments from three to six clusters. To overcome the problem of convergence to local optima, five 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 number of segments that minimize the three specified criteria.

Number of Segments	$-2\log L(\mathbf{\Psi})$	Number of parameters	Entropy	AIC	BIC	ICL
3	3842.7	69	125.11	3980.7	4208.3	4458.5
4	2625.8	92	37.21	2809.8	3113.2	3187.7
5	2549.9	115	42.36	2779.9	3159.2	3243.9
6	2510.3	138	53.61	2786.3	3241.5	3348.7

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

BIC and ICL reach a four-segment solution whereas AIC reaches a five-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 four-segment solution.

After assigning each respondent to the segment with highest posterior probability, these were then categorized by their age group and status. The median age was used as a cut point for the two age categories. Interesting differences between the segments emerge when observed frequencies in each segment are compared by these two demographic variables.

Segment	Age Group	Status		Total
		Single	Married	
1	27 years or less	9	21	30
	More than 27 years	5	17	22
	Total	14	38	52
2	27 years or less	8	3	11
	More than 27 years	18	9	27
	Total	26	12	38
3	27 years or less	44	2	46
	More than 27 years	7	18	25
	Total	51	20	71
4	27 years or less	8	5	13
	More than 27 years	6	20	26
	Total	14	25	39

Table 2 - Frequency of respondents assigned to segments by age group and status

The second statistical objective is to define these segments by investigating how respondents in each segment trade off between different package tours. For ease of interpretation of the fitted model, predicted worth was converted to an expected value of the rating. Thus for each segment, the relationship between worth and price was examined for each destination. The price profiles in Figure 1 characterize different patterns in tourist behaviour. Segment 1, which constitutes a higher proportion of married respondents in the younger age group, represents tourists who worth Paris most and Prague least for their holiday destination. For these tourists high-priced package tours have a negative deterrent effect. Segment 2, which contains a higher proportion of single individuals in the older age group, represents tourists who strongly discriminate between different destination package tours. These tourists prefer Paris most and Amsterdam least and exhibit a strong reliance on price as a signal of quality. Segment 3, which comprises a higher proportion of single respondents in the younger age group, represents tourists who hardly discriminate between city destinations. However, these tourists are price sensitive and worth cheap package tours more than expensive ones. Segment 4, which constitutes a higher proportion of married individuals in the older age group, represents tourists who worth Prague most and London least for their holiday destination. These tourists have a slight reliance on price and consider the middling prices as ideal.

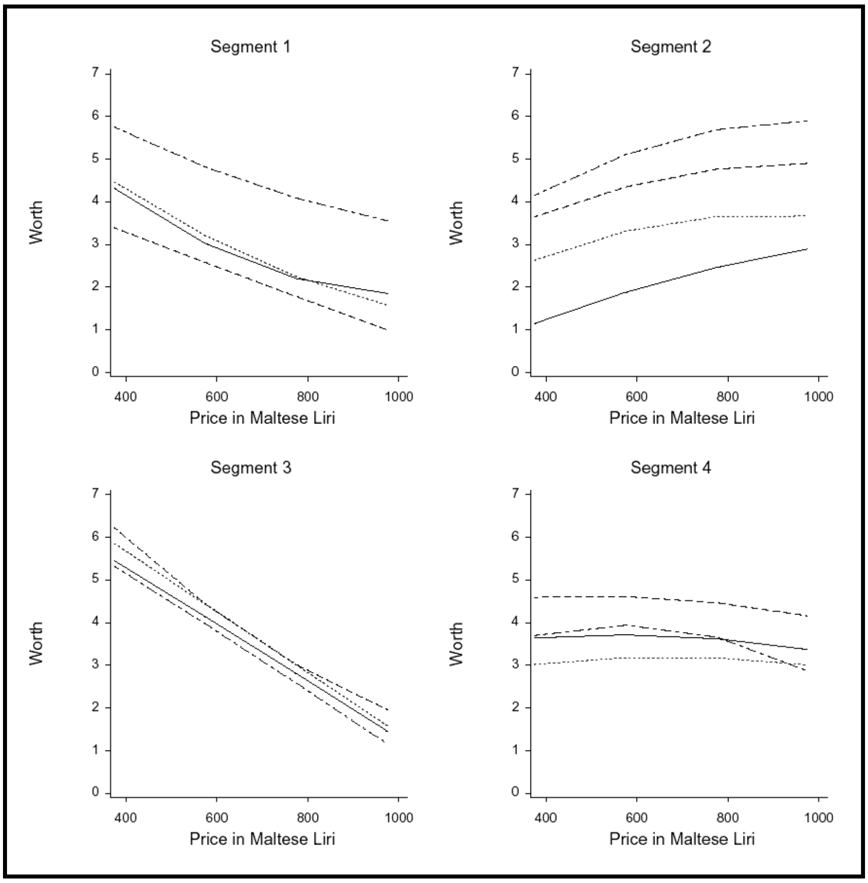

Figure 1 – Price profiles of the expected worth for each destination in the fitted segments

Amsterdam ________________ Paris ___ _ ___ ___ _ ___
London .... Prague ___ ___ ___ ___ ___

4 A Latent Class Choice-based conjoint model

In metric conjoint analysis respondents are asked to provide ratings or rankings to a set of profiles for the response scale, whereas in choice-based conjoint analysis a respondent is asked to choose the most preferred option from several competing alternatives. An advantage of a choice-based task is that it has greater external validity because it mimics what consumers actually do in the marketplace. Another advantage is that a choice-based task is simpler for respondents than rating or ranking alternatives and people will answer choices about almost anything. The major limitation of choice data is that it contains minimal information about consumer preferences. A choice simply indicates the most preferred profile but it does not provide an estimate of the utility of the product profiles. Choice-based conjoint analysis uses the basic ideas and designs of metric conjoint analysis. The procedure that has been developed to identify segments based on choice data is conceptually similar to that for metric conjoint segmentation.

The most popular discrete choice model is the multinomial logit model. The respondent's probability of choosing profile *j*, from a set of *J* profiles, conditional upon belonging to segment *k* is given by:

$$P_k(j) = \frac{\exp\left(\mathbf{x}_j^{'}\boldsymbol{\beta}_k\right)}{\sum_{i=1}^{J}\exp\left(\mathbf{x}_i^{'}\boldsymbol{\beta}_k\right)} \tag{4.1}$$

$n = 1,\ldots,N$ consumers;
$k = 1,\ldots,K$ derived segments;
$j = 1,...,J$ choice conjoint profiles;
π_k is the proportion of respondents in segment *k* and $\boldsymbol{\pi} = (\pi_1,...,\pi_K)$;
y_{nj} is a 0-1 indicator indicating whether the profile *j* is chosen by respondent *n*;
$\mathbf{x}_j$ is the vector containing the values of the explanatory variables corresponding to profile *j*;
$\boldsymbol{\beta}_k$ is the vector of parameter estimates for segment *k* and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_K)^{'}$;

The unconditional choice probability that a respondent chooses profile *j* can be expressed as:

$$P(j) = \sum_{k=1}^{K}\pi_k P_k(j) \tag{4.2}$$

The modelling framework thus entails a finite mixture of conditional multinomial logit models to estimate latent segments with choice data. Given that the choice responses of the *N* respondents are independent, the likelihood function is given by:

$$L(\boldsymbol{\pi},\mathbf{X},\boldsymbol{\beta}) = \prod_{n=1}^{N}\sum_{k=1}^{K}\pi_k\prod_{j=1}^{J}\left[P_s(j)\right]^{y_{nj}} \tag{4.3}$$

The maximization procedure is again carried out using the EM algorithm. Once the parameters $\hat{\pi}_k$ and $\hat{\boldsymbol{\beta}}_k$ are estimated for any iteration of the maximum likelihood procedure, one can assign any consumer *n* to the segment *k* with the highest posterior probability $\hat{p}_{nk}$.

$$\hat{p}_{nk} = \frac{\hat{\pi}_k\prod_{j=1}^{J}\left[\hat{P}_k(j)\right]^{y_{nj}}}{\sum_{k=1}^{K}\hat{\pi}_k\prod_{j=1}^{J}\left[\hat{P}_k(j)\right]^{y_{nj}}} \tag{4.4}$$

5. An Illustrative Application for Choice Data

To illustrate the procedure a further task was included in the study to assess how well the Latent Class choice-based model predicted the respondents' choice behaviour. All the respondents were presented with four choice cards (package tour descriptions) which are displayed in table 3 and were asked to indicate the one they prefered most.

Choice card	Destination	Lodging	Price
A	Amsterdam	Single room	Lm 575
B	London	Single room	Lm 675
C	Paris	Twin room	Lm 475
D	Prague	Twin room	Lm 375

Table 3 – Choice Cards

The observed frequencies in table 4 are obtained by first assigning respondents to the segment with highest posterior probability and then counting their first preferences. The expected frequencies in table 5 are the corresponding total of the predicted choice probabilities given by Eq. (4.1).

	Segment 1	Segment 2	Segment 3	Segment 4	Total
Choice card A	4	11	2	4	21
Choice card B	2	7	2	7	18
Choice card C	29	24	17	11	81
Choice card D	19	8	48	5	80

Table 4 – Observed frequencies of first-choice preferences by choice cards and segments

	Segment 1	Segment 2	Segment 3	Segment 4	Total
Choice card A	1.82	11.2	1.04	3.56	17.6
Choice card B	1.12	13.4	0.25	3.35	18.1
Choice card C	28.4	19.1	14.3	11.9	73.7
Choice card D	22.7	6.36	53.4	8.16	90.6

Table 5 – Expected frequencies of first-choice preferences by choice cards and segments

Visual comparison of the observed and expected frequencies in the four segments shows that the model is picking up the main features of individual choice preferences. The latent class model is correctly eliciting a higher proportion of respondents in segments 1, 2 and 4 who prefer the package tour described by choice card C and a higher proportion of respondents in segment 3 who prefer the package tour corresponding to choice card D. In addition, the model correctly picks out the two most preferred package tours described by the cards C and D.

The following cross-tabulations compare the observed and expected frequencies of first-choice preferences by choice cards, gender, marital status and age groups.

	Males	Females	Total
Choice card A	13	8	21
Choice card B	14	4	18
Choice card C	34	47	81
Choice card D	40	40	80

	Males	Females	Total
Choice card A	9.52	8.05	17.6
Choice card B	10.3	7.90	18.2
Choice card C	37.3	36.3	73.6
Choice card D	43.9	46.7	90.6

Table 6 – Observed and Expected frequencies of first-choice preferences by choice cards and gender

	Single	Married	Total
Choice card A	13	8	21
Choice card B	5	13	18
Choice card C	45	36	81
Choice card D	42	38	80

	Single	Married	Total
Choice card A	7.20	10.4	17.6
Choice card B	7.70	10.5	18.2
Choice card C	39.0	34.6	73.6
Choice card D	51.1	39.5	90.6

Table 7 – Observed and Expected frequencies of first-choice preferences by choice cards and status

	Young	Old	Total
Choice card A	11	10	21
Choice card B	4	14	18
Choice card C	43	38	81
Choice card D	42	38	80

	Young	Old	Total
Choice card A	6.8	10.8	17.6
Choice card B	7.2	11.0	18.2
Choice card C	34.7	38.9	73.6
Choice card D	51.3	39.3	90.6

Table 8 – Observed and Expected frequencies of first-choice preferences by choice cards and age-group

The median age was used as the cut-point between the two age categories. Widows and separated/divorced respondents were assumed to be single. The model correctly elicits a higher proportion of respondents who prefer twin rooms rather than single rooms. This is attributed to the fact that very often tourists prefer to travel with friends or partners with the added benefit that they have to pay less. The model correctly picks very little gender, status and age bias for first choice preferences. However, the model incorrectly picks a higher proportion of respondents who prefer the package tour described by card D rather than the one described by card C. The model correctly elicits low proportions of respondents who prefer the package tours described by cards A and B.

6. Factors affecting the performance of Latent Class metric models (Monte Carlo study)

This Monte Carlo study was conducted to examine the factors that affect the performance of latent class metric conjoint models with respect to parameter recovery, segment membership recovery and computational effort. The simulation was devised to mimic the application in which four destination cities; four price values and two types of hotel rooms were generated to define the product attributes. The price values and the design were set the same as in the application. Two blocks were generated each consisting of 16 distinct attribute level combinations. The two blocks guaranteed a full factorial design and hypothetical subjects were alternately assigned to block. To generate the gender, status and age of these subjects, three sets of N uniformly distributed pseudo-random real values in the range [0, 1] were used. Pseudo random values less than 0.5 corresponded to male subjects in the first set and single subjects in the second set. The generated ages ranged from 15 to 75 years by using a linear transformation of the pseudo random values in the third set. This gives random individual characteristics to each subject. To allocate subjects to segments the proportion π_k of members in each segment was specified, satisfying the constraint $\sum_{k=1}^{K} \pi_k = 1$. This was carried out by first generating N uniformly distributed pseudo-random real values in the range [0, 1] and then by computing the cumulative probabilities $q_k = \sum_{i=1}^{k} \pi_i$. Every subject whose corresponding value was in the range $\left[q_{k-1}, q_k\right]$ was allocated to segment k. This gives a random segment allocation to each hypothetical subject. To simulate the subjects' rating responses, the linear predictors and the corresponding parameters $\boldsymbol{\beta}_k$ were specified for the K segments. Given the segment membership allocation, 16 synthetic data values were generated for each subject. These values were then perturbed by adding an error term having a normal distribution. Six specified cut-points α_r were used to convert these values to rates ranging from 1 to 7. Values in the range (α_{r-1}, α_r) were converted to rate r. This gives a random worth category allocation for each hypothetical response. 120 synthetic datasets were generated according to the following three factors.

1. Number of hypothetical subjects: $N = 100$ or $N = 200$
2. Number of segments: $K = 2$ or $K = 4$
3. Number of predictors: $S = 3$ or $S = 6$

These three factors were chosen because they are expected to affect the performance of the algorithm. The factor levels were selected to reflect a variation in conditions representative of practical applications. In the construction of synthetic data two linear predictors were used. The first included product-related attributes only $S = 3$ and the second included both product-related and individual attributes $S = 6$. 15 data sets were generated for each of the eight factor level combinations and each data set was fitted to the latent class metric conjoint model, presented in section 2. To overcome problems related with convergence to local maxima five starting values were considered for each fit. These were selected from a wide range of seed numbers and the solution with the smallest deviance was selected. The estimated parameters were then compared to the true parameters using the following measures to assess parameter and segment membership recovery.

1. The percentage of correctly classified subjects into their true segments on the basis of the estimated segment membership.

2. The statistic $RMS(\hat{\boldsymbol{\beta}})$ is used to assess recovery of the P parameters.

$$RMS(\hat{\boldsymbol{\beta}}) = \left[\sum_{p=1}^{P} \frac{(\beta_p - \hat{\beta}_p)^2}{P} \right]^{\frac{1}{2}} \tag{6.1}$$

3. The statistic $RMS(\hat{\boldsymbol{\pi}})$ is used to assess recovery of the K segment proportions.

$$RMS(\hat{\boldsymbol{\pi}}) = \left[\sum_{k=1}^{K} \frac{(\pi_k - \hat{\pi}_k)^2}{K} \right]^{\frac{1}{2}} \tag{6.2}$$

The mean performance measures for parameter recovery were computed by averaging the values of $RMS(\hat{\boldsymbol{\beta}})$ and $RMS(\hat{\boldsymbol{\pi}})$ across the 15 data sets of a particular factor level combination. Similarly, the mean performance measure for segment membership recovery was computed by averaging the percentages of correctly classified subjects across the 15 data sets.

7. Results

Table 9 shows the mean performance measures for parameter and segment membership recovery at each factor level combination. The mean $RMS(\hat{\boldsymbol{\pi}})$ is sensitive to changes in the number of subjects but less sensitive to changes in the number of segments. Recovery of segment membership proportions improve with an increase in the number of subjects and number of segments but are hardly affected by the number of predictors. The mean $RMS(\hat{\boldsymbol{\beta}})$ is sensitive to changes in the number of subjects, predictors and segments. Parameter recovery improves with an increase in the number of subjects and number of predictors but deteriorates with an increase in the number of segments. Segment membership recovery decreases with an increase in the number of segments but is hardly affected by the number of subjects and predictors. The computational effort was measured by recording the number of iterations required for convergence at each fit. Computational effort is affected mostly by a change in the number of segments but to a lesser extent by the number of subjects. Increasing the number of segments increases the number of iterations required.

Number of segments	Number of subjects	No of predictors	$RMS(\hat{\boldsymbol{\beta}})$	$RMS(\hat{\boldsymbol{\pi}})$	Segment Recovery
2	100	3	0.9459	0.0406	99.61%
		6	0.6338	0.0443	99.82%
	200	3	0.8724	0.0291	99.65%
		6	0.5563	0.0361	99.91%
4	100	3	1.1301	0.0385	98.47%
		6	0.9442	0.0416	98.93%
	200	3	0.9664	0.0273	98.57%
		6	0.8666	0.0249	98.97%

Table 9 - Mean performance measures for parameter and segment membership recovery

References

[1] Banfield, C.F., Bassil, L.C.: *A Transfer Algorithm for Nonhierarchical Classification.* In Applied Statistics, Vol. 26, pp. 206-210, 1977.

[2] Biernacki, C., Govaert, G.: *Using the Classification Likelihood to Choose the Number of Clusters*. In Computing Science and Statistics, Vol. 29, pp. 451-457, 1997.

[3] Camilleri, L., Green, M.: *Statistical Models for Market Segmentation*. In Proceedings of the 19th International Workshop on Statistical Modelling, Florence, pp. 120-124, 2004.

[4] Celeux, G., Soromenho, G.: *An Entropy Criterion for assessing the number of Clusters in a Mixture model*. In Classification Journal, Vol.13, pp. 195-212, 1996.

[5] Dempster, A.P., Laird, N.M., and Rubin, D.B.: *Maximum Likelihood from Incomplete Data via the EM algorithm*. In Journal of the Royal Statistical Society, Vol. 39, pp. 1-38, 1977.

[6] DeSarbo, W.S., Oliver, R.L., Rangaswamy, A.: *A Simulated Annealing Methodology for Clusterwise Linear Regression*, In Psychometrika, Vol. 54, 707-736, 1989.

[7] DeSarbo, W.S.,Ramaswamy, V., Cohen, S.: *Market Segmentation with Choice-Based Conjoint Analysis,* In Marketing Letter 6:2, pp. 137-147, 1995.

[8] DeSarbo, W.S., Wedel, M., Vriens, M., Ramaswamy, V.: *Latent Class Metric Conjoint Analysis*, In Marketing Letters, Vol. 3, pp. 273-288, 1992.
[9] Elrod, T., Louviere, J.J., Davey, K.S.: *Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models*, In Journal of Marketing Research, Vol. 29, pp. 368-377, 1992.
[10] Francis, B., Green, M. and Payne, C. (1993), *The GLIM 4 manual*, Oxford Science Publications.
[11] Green, M.: *Statistical Models for Conjoint Analysis*, In Proceedings of the 15th International Workshop on Statistical Modelling, Bilbao, pp. 216-222, 2000.
[12] Green, P.E., Krieger, A.M.: *Segmenting Markets with Conjoint Analysis*, In Journal of Marketing, Vol. 55, pp. 20-31, 1991.
[13] Green, P.E., Krieger, A.M., Wind, Y.: *Thirty Years of Conjoint Analysis: Reflections and Prospects*, In Interfaces, Vol. 31, Part2, pp. 56-73, 2001.
[14] Green, P.E., Rao, V.R.: *Conjoint Measurement for Quantifying Judgemental Data*, In Journal of Marketing Research, Vol. 8, pp. 355-363, 1971.
[15] Green, P.E., Srinivasan, V.: *Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice*, In Journal of Marketing, Vol. 54, pp. 3-19, 1990.
[16] Gustafsson, A., Herrmann, A. and Huber, F.: *Conjoint Analysis as an Instrument of Market Research Practice*, In Conjoint Measurement: Methods and Applications, Gustafsson, A., Herrmann, A. and Huber, F. (ed), pp. 5-46, Springer, 2000.
[17] Haaijer, R., Wedel, M.: *Conjoint Choice Data: General Characteristics and Alternative Model Specifications*, In Conjoint Measurement: Methods and Applications, Gustafsson, A., Herrmann, A. and Huber, F. (ed), pp. 345-386, Springer, 2000.
[18] Hagerty, M.R.: *Improving the predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis*, In Journal of Marketing Research, Vol. 22, pp. 168-184, 1985.
[19] Kamakura, W.A.: *A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments*, In Journal of Marketing Research, Vol. 25, pp. 157-167, 1988.
[20] McLachlan, G. and Peel, D.: *Finite Mixture Models*, Wiley Series in Probability and Statistics, 2000.
[21] Ogawa, K.: *An Approach to Simultaneous Estimation and Segmentation in Conjoint Analysis*, In Marketing Science, Vol. 6, pp. 66-81, 1987.
[22] Ramaswamy, V., Cohen, S.H.: *Latent Class Models for Conjoint Analysis*, In Conjoint Measurement: Methods and Applications, Gustafsson, A., Hermann, A. and Huber, F. (ed), pp. 415-446, Springer, 2000.
[23] Spath, H.: *Clusterwise Linear Regression*, In Computing, Vol. 22, pp. 367-373, 1979.
[24] Vriens, M., Wedel, M., Wilms, T.: *Metric Conjoint Segmentation Methods A Monte Carlo Comparison*, In Journal of Marketing Research, Vol. 23, pp. 73-85, 1996.
[25] Wedel, M., DeSarbo, W.S.: *A Mixture Likelihood Approach for Generalized Linear Models*, Journal of Classification, Vol. 12, pp. 1-35, 1995.
[26] Wedel, M., DeSarbo, W.S.: *A Review if Recent Developments in Latent Class Regression Models*, In Advanced Methods of Marketing Research, Bagozzi R.P. (ed), pp. 352-388. Cambridge, M.A: Blackwell.
[27] Wedel, M., Kamakura, W.A.: *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, 2000.
[28] Wedel, M., Kistemaker, C.: *Consumer Benefit Segmentation using Clusterwise Linear Regression*, In International Journal of Research Marketing, Vol. 6, pp. 45-49, 1989.
[29] Wedel, M., Steenkamp, J.B.: *Fuzzy Clusterwise Regression Approach to Segmentation*, In International Journal of Research in Marketing, Vol. 6, pp. 241-258, 1989.

[30] Wedel, M. and Steenkamp, J.B.: *A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation*, In Journal of Market Research, Vol.28, pp. 385-396, 1991.

[31] Wittink, D.R., Cattin, P.: *Commercial Use of Conjoint Analysis*, In Journal of Marketing, Vol. 53, pp. 91-96, 1989.

[32] Wittink, D.R., Vriens, M., Burhenne, W.: *Commercial Use of Conjoint Analysis in Europe*, In International Journal of Research in Marketing, Vol. 11, pp. 41-52, 1994.