Comparison of Different Clustering Methods for Investigating Individual Differences using Choice Experiments

Asioli, D.[1,2,3], Berget, I.[2], Næs, T.[2,4]

[1]University of Reading, Reading, United Kingdom.

[2]Nofima AS, Ås, Norway.

[3]University of Arkansas, Fayetteville, United States.

[4]University of Copenhagen, Copenhagen, Denmark.

ABSTRACT

Different strategies for investigating individual differences among consumers using choice experiments are compared. The paper is based on a consumer study of iced coffee in Norway. Consumers (n = 102) performed a choice task of twenty different iced coffee profiles varying in coffee type, production origin, calorie content and price following an orthogonal design. Consumer factors, such as socio-demographics, attitudes and habits, were also collected. Choice data will be analysed using two different clustering strategies. Strategy one is the most classical approach called Latent Class Logit (LCL) model, while Strategy two uses Mixed Logit (ML) model combined with Principal Component Analysis (PCA) for visual segmentation or with automatic clustering detection using Fuzzy C Means clustering (FCM). The clusters obtained can be interpreted using external consumer factors by using the Partial Least Square – Discrimination Analysis (PLS-DA) model. The different approaches are compared in terms of data analysis methodologies, modeling, outcomes, interpretation, flexibility, practical issues and user friendliness.

KEY WORDS: iced coffee, consumers, choice experiment, clustering methods, method comparison, Norway.

1. INTRODUCTION

Choice experiments (CEs) have been developed for investigating consumers' choices both for market and non-market goods (Louviere, Hensher, & Swait, 2000). Based on an experimental design, different product alternatives are created using different combinations of attribute levels that characterize the goods (Louviere et al. 2000). Then, consumers are presented with a series of alternative product choice scenarios (i.e. choice sets) that mimic real market scenarios, and are asked to choose their most preferred option (i.e. product alternative) within each choice scenario. One of the arguments put forward for CEs is that having respondents choose a single preferred stimulus among a set of stimuli better approximates a real purchase situation (Carson et al., 1994; Louviere et al., 2000).

In 1956 Smith defined: "Market segmentation involves viewing a heterogeneous market as a number of smaller homogeneous markets, in response to differing preferences, attributable to the desire of customers for more precise satisfaction of their varying wants". Thus, market segmentation captures the natural phenomena of consumer heterogeneity by allowing to distinguish homogeneous groups of customers who can be targeted in a similar way because they have similar needs and preferences (Allenby, Arora, & Ginter, 1998; Wedel & Kamakura, 2002). Combris, Bazoche, Giraud-Héraud, & Issanchou (2009) describe consumer heterogeneity as "a key and permanent feature of food choice". It is today an important and natural element of food choice research (Frewer & van Trijp, 2007; MacFie, 2007). Preference heterogeneity can be investigated in terms of consumer factors (i.e. demographics, attitudes, habits, etc.).

There exist different ways of studying consumer hetergoneity, either by studying individual differences per se as done in for instance preference mapping, or by segmentation. The latter can be done both *a priori* and *a posteriori* as was discussed in for instance (Næs, Kubberød,

& Sivertsen, 2001; Næs, Brockhoff, & Tomic, 2010). The *a priori* segmentation is based on splitting the consumer group into subgroups according to consumer factors, and then analyzing the group preferences separately or together in an ANOVA model or a Mixed Logit (ML) model (depending on data collection, see e.g. Asioli, Næs, Øvrum, & Almli, 2016a). A *posteriori* segmentation, on the other hand, is based on creating consumer groups of similar product preferences by analyzing the actual preferences, liking or purchase intent data to create segments, and then afterwards relating segments to consumer characteristics *a posteriori.* The most frequently used method for *a posteriori* segmentation in CE is based on the Latent Class Logit (LCL) model (Greene & Hensher, 2003), but other possibilities also exist as will be discussed below.

The main objective of this study is to compare the standard and established LCL method with an alternative approach based on clustering of the regression cofficients from a Mixed Logit (ML) model (Asioli, Almli, et al. (2016b). This approach offers some potential advantages both from an interpretation and flexibility point of view. Most importantly, the ML approach suggested offers better opportunities for visual interpretation of the results. In addition, since it is not scale invariant, it opens up the possibility for a more explorative investigation of cluster structure. Data analysis methodologies, outcomes, interpretation, practical issues and user friendliness of the two strategies will be discussed. To compare the methods, data from a CE investigating consumers' preferences for iced coffee products in Norway were used. For interpreting the segments in terms of external consumer factors we will here use the Partial Least Squares – Discriminant Analysis (PLS-DA) model (Asioli, Almli, & Næs, 2016b; Ståhle & Wold, 1987).

2. THEORY: STATISTICAL METHODS USED

2.1 Utility and choice models

Choice-based data are routinely analysed within a utility framework by so-called discrete choice models (DCMs) (Louviere et al., 2000; Train, 2009). The approach is based on modelling "Utility", that is to say the net benefit a consumer obtains from selecting a specific product in a choice situation, as a function of the conjoint factors. The utility of a product j for individual n in a choice occasion t (choice set) is written:

$$U_{njt} = \boldsymbol{\beta}'_n \mathbf{x}_{jt} + \varepsilon_{njt} \tag{1}$$

where $\boldsymbol{\beta}_n$ is a vector of individual-specific parameters accounting for preference heterogeneity, $\mathbf{x}_{jt}$ is a vector of conjoint factors, and ε_{njt} is a random error term. Under the assumption that the random errors follow a so-called extreme value distribution (Train, 2009) and are independent and identically distributed (i.i.d) the choice probability for product j and choice set t given the parameter $\boldsymbol{\beta}_n$ has a simple form:

$$P_{njt} = \frac{\exp(\boldsymbol{\beta}'_n \mathbf{x}_{jt})}{\sum_{i=1}^{J_t} \exp(\boldsymbol{\beta}'_n \mathbf{x}_{it})} \tag{2}$$

where J_t is the number of products in choice set t. This function will play a central role in both strategies used in this paper.

Among the different DCMs that are able to capture consumers' heterogeneity, we will here focus on two of the choice models most applied: the Latent Class Logit (LCL) model and the Mixed Logit (ML) model (Greene & Hensher, 2003; Train, 2009). The LCL model assumes constant model parameters within each cluster and captures consumer heterogeneity assuming a mixing distribution for the clusters (Greene & Hensher, 2003; Hess, Ben-Akiva, Gopinath, & Walker, 2011). The ML model, on the other hand, captures individual consumer

5

heterogeneity by assuming random regression parameters that follow a (for instance the normal) distribution.

2.2 STRATEGY ONE: Investigation of individual differences using the Latent Class Logit (LCL) model

The LCL model assumes that the consumer group can be split in subgroups with a constant β vector in each group (Greene & Hensher, 2003). In other words, the LCL model assumes that a discrete number of classes are sufficient to account for preference heterogeneity across classes (Shen, 2009).

The choice probability that an individual of class s chooses alternative j from a particular set constituted of J_t alternatives, is expressed as:

$$P_{j/s} = \frac{\exp(\beta_s' \mathbf{x}_{jt})}{\sum_{i=1}^{J_t} \exp(\beta_s' \mathbf{x}_{it})} \tag{3}$$

where $s = 1,...S$ represents the number of classes and β'_s is the fixed (constant) parameter vector associated with class s. In order to establish the likelihood, these choice probabilities have to be multiplied over the choice sets and finally combined over all individuals.

The LCL model is widely applied in CEs in different fields such as transportation, environment, marketing, health and also food marketing (Caputo, Nayga, & Scarpa, 2013; Ortega, Wang, Wu, & Olynk, 2011; Peschel, Grebitus, Steiner, & Veeman, 2016).

To estimate the LCL model it is possible to use the Expectation – Maximization (EM) algorithm which allows for a good numerical stability and good performance in terms of run time (Bhat, 1997; Pacifico & Hong il, 2013; Train, 2008). One of the main issues in the LCL model is the choice of S, which is the number of latent classes. Given the fact that S is not a

parameter, it is not possible to test it directly (Shen, 2009). Louviere et al. (2000) suggested a number of methods based on Akaike Information Criterion (AIC) and Consistent Akaike Information Criterion (CAIC) to guide the model selection and they suggest that the model that minimizes AIC and CAIC should be prefered (see for details Louviere et al. 2000). It should, however, be mentioned that since consumer studies of this type seldom present clearly separated clusters (Endrizzi, Gasperi, Rødbotten, & Næs, 2014) selecting the number of clusters can in many cases also be a matter of choice related to, for instance keeping the size of the clusters reasonably large.

In this paper, the Latent Class Logit (LCL) model used will include both main effects and two-way interactions in order to calculate the class parameters β_s. The main results from the methods are the subgroups, the regression parameter within each group and indications of how well each consumer fits to the different subgroups. A posteriori probabilities of membership are also provided and will be discussd briefly below. The method is invariant to the relative scale of the input variables.

2.3 STRATEGY TWO: Investigation of individual differences using the Mixed Logit (ML) model

2.3.1   Mixed Logit Model approach

Among the different DCMs, one of the more frequently applied is the Mixed Logit (ML) model due to its flexibility (Train, 2009). In the ML model, the utility is defined as above with the exception that now, the β is assumed to be a random vector (i.e. each individual will have their own and unique value). The choice probabilities are then combined giving a likelihood that can be maximized/optimized by simulation methods (Train, 2009). An advantage of the ML model is that one may freely include random parameters β_n of any distributions and

correlations between random factors which allows for models that better match real-world situations (Train, 2009).

By using PCA on the regression coefficients, organised with consumers as rows and regression coefficients as columns, it is possible to analyse the individual differences *per se* by using scores plots and loadings plots directly for interpretation without any focus on segmentation. In the following we will, however, for the purpose of comparison with the LCL method, use an automatic segmentation procedure, and use the PCA scores for illustration of clusters.

Clustering of regression coefficients

Automatic clustering can be performed by hierarchical or by partitioning methods. Each has different advantages, but in some studies the partitioning methods perform better than hierarchical methods (see Wajrock, Antille, Rytz, Pineau, & Hager, 2008). More specifically we suggest using the Fuzzy C Means (FCM) algorithm (Bezdek, 1981). FCM is a generalization of the well known K means methodology, but instead of providing a crisp partition of the data, where each observation belongs to one segment only, fuzzy clustering give a more quantitative description of clusters by estimating membership values that describe how well each consumer belongs to each cluster. The FCM has good convergence properties and is flexible with respect to distance measure used. The membership values make fuzzy clustering methods useful in situations with no clear separation between clusters. Note that for the purpose of the comparison done here, other methods than the FCM could have been selected and no special emphasis will be given to the specifics of the fuzziness concept. For further properties and advantages of the method we refer to for instance Bezdek (1981) and Berget, Mevik, & Næs (2008).

The general criterion to be minimized by FCM can be written as

$$L = \sum_{i=1}^{N} \sum_{s=1}^{S} v_{is}^{m} d_{is}^{2} \qquad (4)$$

Here the v's represent membership values (within the interval 0 to 1, and must sum to one for each object, here consumer) and m is the fuzzifier parameter, usually set equal to m=2 (Bezdek, 1981). In ordinary FCM, the d is the Euclidean distance between observations and the cluster centre, but many other possibilities exist. The sum is taken over S clusters, and N objects (consumers). The criterion is minimized by an iterative procedure that computes cluster centres for given membership values and vice versa (Bezdek, 1981). The procedure is initialized randomly, or according to previous knowledge.

As for the LCL model the number of clusters has to be decided in advance. Typically, partition methods, such as FCM, are run for a range of S (number of clusters), the number of clusters to use is then determined from a combination of cluster validation parameters (Halkidi, Batistakis, & Vazirgiannis, 2001), interpretational aspects and practical considerations. There exist a number of indices which can be applied for selecting the number of clusters, here we applied Normalised Hubert gamma, the Xie-Beni and the Fukuyama-Sugeno indeces as well as the Fuzzy Hypervolume (Halkidi, Batistakis, & Vazirgiannis, 2001).

Clustering is a difficult exercise, and should not always be considered a one-shot procedure. Interpretation as well an explorative approach may in some cases be important. A possible advantage of the regression coefficient approach is that clustering can be performed not only on the regression coefficients as they are, but also on coefficients that are standardised or transformed in other ways based on prior knowledge about what aspects that should be emphasized. This opens up the opportunity for highlighting aspects that are less visible, for instance coefficients with less variability and also for eliminating coefficients, i.e variables, from the clustering. This means that this approach opens for a more explorative way of

analyzing the data, looking at differenet options and weighting schemes. This opportunity is not shared by the scale invariant LCL approach and will be considered further below.

3. MATERIALS AND METHODS

3.1 Consumer test

In order to test the two strategies presented in the Section 2, we used a dataset based on iced coffee products. 102 consumers were recruited in the region south of Oslo, Norway, in Autumn 2012 using a survey based on a CE. The test included different sessions, one of them being a choice task. The experiment is the same as used in Asioli et al., (2016b), but for completeness a short review is given below.

3.2 Iced coffee products

The iced coffee products were artificially created using four factors identified based on focus group results. The factors were coffee type, calorie content, origin and price; see Asioli, Næs, Granli, & Lengard Almli (2014). All the factors have two levels, except price with three levels (Table 1).

Table 1 – Conjoint factors, levels and codes

<<Please, place here table 1>>

3.3 Choice task

10

The product profiles were created based on the combinations of factors and levels described in Section 3.2. We used an orthogonal choice design composed of eight choice sets of three products. The design was generated in SAS version 9.3. The design featured 20 unique samples where all of them were taken from the full factorial design (see Asioli, Næs, et al., 2016a for more details).

The eight triads of iced coffee profiles were displayed successively on a computer screen in the form of photographs. Product presentation was randomized across participants both at choice set level, and at product level within choice sets. For each choice-set, consumers' probability of buying was elicited with the question: *"Imagine that you are purchasing iced coffee. Which of these iced coffees are you most likely to buy?"* and participants answered by clicking on one of the three alternatives.

3.4 Consumer factors

In addition to the choice task, a number of consumer factors were collected. The factors investigated are related to warm and iced coffee consumption habits, food attitudes such as food neophobia, health consciousness and ethnocentricity, and socio-demographics.

Consumer factors are measured using both numerical and categorical variables. For the importance of attributes for choosing iced coffee, the scale is anchored in 1 (Not important at all) and 5 (Very important at all). The same is the case for the habits factors. All the categorical attributes have been coded as a number of dummy variables where 0 represents the absence of the actual level while 1 represents the presence of the attribute level. In the example below, the significant attributes will be presented (Table 3). The complete list of consumer factors can be obtained from the authors.

3.5 Data analysis

All two-level conjoint factors were coded using effects coding (-1; 1) (Bech & Gyrd-Hansen, 2005), and price was coded in three levels (mean centered) (-1; 0; 1) (see Table 1). In other words, the price was coded as a linear covariate as this allowed to handle its three experimental levels in a simple manner and was considered appropriate as previous modelling on these data have indicated a linear effect of price (Asioli et al., 2014; Asioli et al., 2016a).

3.5.1 STRATEGY ONE: Investigation of individual differences using the Latent Class Logit (LCL) model

In the model both main effects and interactions are presented. Interaction variables are obtained by multiplication of the main effects variables. The utility LCL model for iced coffee *j* in choice occasion *t* and cluster *s* is written:

$$U_{jt} = \beta_1\ Coffee_{jt} + \beta_2\ Calories_{jt} + \beta_3\ Origin_{jt} + \beta_4\ Price_{jt} + B_5\ (Coffee* Calories)_{jt} +$$
$$\beta_6\ (Coffee*Origin)_{jt} + \beta_7\ (Coffee*Price)_{jt} + \beta_8\ (Calories*Origin)_{jt} + \beta_9$$
$$(Calories*Price)_{jt} + \beta_{10}\ (Origin*Price)_{jt} + \varepsilon_{jt} \qquad (4)$$

The cluster index *s* is omitted from the regression coefficients for simplicity. Following Pacifico & Hong il (2013) the LCL model based on this utility was estimated using the module *lclogit* in STATA 11.2 software (StataCorp LP, College Station, US) in two steps. First, we identified the optimal number of latent classes by examining the Bayesian Information Criteria (BIC) and CAIC. Second, we entered the optimal number of clusters

identified in the first step into the LCL model and estimated it. The regression coefficients for each class as well as their significance level will be reported.

The choice probabilities from the LCL model provide information about degree of membership for consumers to the groups, and are comparable to the membership values from the FCM approach. Neither choice probabilities from LCL nor membership values from FCM will be highlighted here since they do not provide additional arguments for the comparison below.

3.5.2 *STRATEGY TWO: Investigation of individual differences using the Mixed Logit (ML) model*

As for Strategy one, we included all the main effects and their two-way interaction, i.e. the same model as above. Then, the matrix of individual parameter estimates $\mathbf{\beta}_n$ was extracted from the ML model by using the command *mixlbeta* in STATA 11.2. The coefficients are assumed to be independently normal distributed.

The matrix of individual parameter estimates $\mathbf{\beta}_n$ extracted from the ML model analysis (organized with consumers as rows) is submitted to Principal Component Analysis (PCA) in order to identify and interpret the main components of variation between individuals. PCA was conducted in the multivariate statistical software package The Unscrambler X 10.3 (Camo Software AS, Norway).

The FCM algorithm was applied using in-house programs in Matlab (Mathworks Inc). The estimated coefficients from the ML model were used as input to the method. Each column in

the data matrix (i.e. for each regression coefficient) was mean centered. This centering does not affect the clustering, and was done to facilitate interpretation. Since some of the regression coefficients have a much larger variance than the others, an approach based on standardized coefficients was also attempted in order to illustrate the flexibility and possible advantage of this approach as discussd above. The fuzzifier parameter m was set equal to two in all cases, and the algorithm was run with two to six clusters. As partition methods such as FCM can be sensitive to the initialization, the algoritm was run with ten initialisations for each value of S (S=2,3,..,6).

3.5.3 Investigation of consumer factors using PLS-DA

In this paper, PLS-DA models were run on standardized input variables, using cross-validation on 10 random segments and performing a jack-knife uncertainty test with 95% confidence interval for the detection of significant variables (Martens & Martens, 2000). We used a two-step procedure: in the first step all the consumer factors were included in the model. Then, in the second step a new PLS-DA model was run only including significant consumer factors from the first step. For the PLS-DA the consumer groups were represented by dummy variables (Ys) in the PLS-DA, while consumer factors were used as independent variables (Xs). Calculations were performed in The Unscrambler X 10.3 (Camo Software AS, Oslo).

3.6 Methods comparison

The comparison will be based on interpretation of regression coefficients, PCA plots and interpretability in terms of relations to consumer factors. Although the focus of this paper is

14

on principles and procedures, alse the actual clusters themselves for the different methods will be compared.

4. RESULTS

4.1 STRATEGY ONE: Investigation of individual differences using the Latent Class Logit (LCL) model

According to Nylund, Asparouhov, & Muthén (2007) the optimal number of clusters for the LCL model is when CAIC and BIC parameters are minimized. Thus, we run different LCL models and then calculated the CAIC and BIC parameters respectively for each model. CAIC are 1290.5, 1267.5, 1291,9, 1301.4, 1372.8 respectively for 2,3,4,5 and 6 cluster solutions while BIC are 1268.5, 1233.5, 1245.9, 1243.4, 1302.8 respectively for 2,3,4,5 and 6 cluster solutions. Thus the optimal number of clusters for LCL was found to be 3. However, the LCL model did not converge for any of the starting values or convergence criteria used. We therefore decided to consider both a 2-cluster solution which converged nicely and a 3-cluster solution after a large number of iterations, but without a proper convergence. Two of the clusters in the latter solution were very similar to the 2 clusters in the 2-cluster solution. None of the factors for the third cluster were significant indicating a situation with no clear separation among the clusters. This interpretation will be discussed further below.

The results of the LCL model with 2-clusters solution are reported in Table 2. For comparison we have also included the mean and standard deviation results for the regression coefficients from the ML model to be discussed below (see Section 4.2).

Table 2 - Estimated parameters for LCL and ML models with conjoint variables' main effects and interactions. The two columns to the left correspond to the estimated

regression coefficients for the two cluster based on LCL model while the two columns to the right refer to the population effects based on the ML model (mean and SD).

<<Please, place here table 2>>

Consumers belonging to group 1 (Espresso group - 49 consumers) have strong preferences for the espresso and low price iced coffee products. We can notice that only one interaction is significant (Coffee*Price). Consumers who prefer espresso iced coffee products are a little bit less sensitive to price changes than consumers who prefer latte (see Asioli, Næs, et al., 2016a, for more details). In the group 2 (Latte group - 53 consumers) consumers have strong preference for latte, low calories, Norwegian origin and low priced iced coffee products. As for group 1, the only significant interaction is Coffee*Price meaning that consumers who prefer latte iced coffee products are a bit more sensitive to price changes than consumers who prefer espresso (see Asioli, Næs, et al., 2016a, for more details). The main differences between the two segments is therefore the difference in preference for coffee type and that segment 2 seems to distinguish more the calory levels and the two origins. All main effects except coffee type go in the same direction as functions of the design variables.

The a posteriori probabilities for membership, as obtained in the STATA implimentation, are all very high (or very low), few below 0.99, and were not found to be very useful for interpretation in this case.

4.2 STRATEGY TWO: Analysis of the regression coefficients

4.2.1 *Estimated parameters from the Mixed Logit (ML) model*

Table 2 contains the estimated parameters of the ML model (means and SDs) for the main and interaction effects of the conjoint factors at population level, as well as as the variability of the individual coefficients as measured by SD. The null hypothesis that all coefficients are zero is rejected by a Wald test (p-value <0.01). Looking at the results from ML models, we can again see that on average consumers prefer low calories, low prices and Norwegian origin, while coffee type is not significant at mean population level. In terms of consumers' heterogeneity, the SDs of all main effects have significant values with the larger magnitude for coffee type. This finding combined with the no significance of the regression coefficient for coffee type, is clearly confirmed by the the two estimated regression coefficients for coffee type in the LCL model that are both significant and have high magnitudes with opposite signs. This confirms that the coffee type is the main conjoint variable for describing individual difference although the population main effect for coffee type is close to 0.

The loadings for the two first PCs are given in the Figure 1. The four first components correspond very closely to the four main effects in the study, which is quite natural since the coefficients are assumed independent in the ML model. The components explain, 86%, 6%, 4% and 3% and represent Coffee type, Origin, Calories and Price respectively, corresponding well with the significance results above. Thus, it is clear that Coffee type (i.e. Espresso vs. Latte) explains the largest variance which indicates that consumers have clear preferences for Espresso or Latte iced coffee. One can see from the scores plot (Figure 2) that there is a continuum of individual differences covering the whole space. This shows that there are no clearly separated segments in the data set and therefore segmentation in this case is merely a way of organizing the consumers in subgroups that have a similar preference profile.

Figure 1 – PCA correlation loadings plot for PC-1 and PC-2 on individual ML parameter estimates from choice data. The interactions effect that overlap (i.e. non clearly understandable) in the middle of the plot are: espresso_highcalories, espresso_italy, espresso_highprice, highcalories_italy, highcalories_highprice and italy_highprice.

<<Please, place here figure 1>>

4.2.2. Investigation using Fuzzy C Means (FCM) algorithm

The FCM was run as described in the Section 3.5.2. All performance indices tested (i.e. XieBeni, Fukujama – Sugeno, Fuzzy Hypervolume and Compactness) suggested that 3 (or 4) clusters is the better solution. Figure 2 shows the results for the different indices[1].

Figure 2 - Plots of the four indices tested for cluster number in FCM (non-standardised data)

<<Please, place here figure 2>>

Figure 3 shows the PCA scores plot split into three clusters. As can be seen, the three clusters are identified along the first principal component, with one cluster in the middle and the two others representing a Latte and an Espresso group respectively. This is very natural given the much larger variance of the coffee type coefficient.

---

[11] More details for cluster validation methods are available in Halkidi et al. (2001).

Figure 3 - The results from FCM for three clusters plotted in the PCA scores plot. To ease interpretation, the names "Latte", "Espresso" and "Segment 3" have been superimposed. The three segments are indicated by different symbols (colors), blue squares represent the latte segment, green triangles the espresso segment and red circles the third, middle segment.

<<Please, place here figure 3>>

With two clusters fixed, the splitting based on FCM was almost identical to the one obtained by visual segmentation as done in Asioli, Almli, et al. (2016b). In that paper the two clusters considered were obtained by splitting according to positive or negative score along the first component.

The membership values covered in this case cover a larger range than the a posteriori probabilities from the LCL, and can therefore be used more easily for judging clear of less clear membership.

4.2.3. Interpretation of the three LCL segments using PCA

At this point it is useful to see how the segments from the LCL model relates to the plots above. The two clusters solution was almost identical to the one obtained by FCM and visual interpretation (Asioli, Almli, et al., 2016b). When comparing the three cluster solution for LCL with Figure 3, only 7 consumers (of 102) were clustered differently, hence the two approaches give very similar results. This structure may shed some light on the lack of convergence of the 3 segments solution and the lack of significance of the coefficients; the cluster in the middle simply does not clearly separate the two coffee preference groups and therefore the algorithm does not find any good reason for estimating it. Again a PCA solution helps interpretation of results.

19

4.2.4. Clustering based on standardized variables

Above we argued that clustering should not always be considered a one-shot exercise, but should merely be considered an explorative exercise based on trying different options, possibly based on prior knowledge, with subsequent interpretation, for instance based on comparison with external consumer factors as illustrated below. This may be important for suggesting other and more subtle segments that do not necessarily only depend on the most dominating variable, in this case coffee type, In order to visualize this possibility, we repeated the FCM on coefficient data after standardization of the regression coefficients for the main effect. We emphasise that this is just and illustration of the possibility and how it can be interpreted by the use of external consumer factors, and that other options may be more suitable in other circumstances. Our concrete illustration here corresponds to giving all main effects the same weight in the segmentation procedure instead of highlighting only one component as done above.

The results obtained with FCM and three clusters based on standardisation of the regression coefficients from the main effects are given in Figure 4 for 3 clusters (the number of clusters was more difficult to identify in this case) in the same PCA plot as used above. Figure 4 shows that after standardization, the splitting is different. One cluster (empty squares, called "Latte") is to a large extent overlapping with the latte segment from the previous results, whereas the other two clusters (+ and x, respectively "Norwegian espresso" and "Italian espresso" groups) are split along the second component (i.e. origin). Although the results for this data set were not very surprising, these findings illustrate that different ways of using the coefficients data may give different results and then different proposals for how to segment the data. The results should as always be interpreted in terms of consumer attributes as discussed below.

Figure 4 – Score plot obtained with unstandardized data, and clusters obtained with FCM on standardized data and S=3. To ease interpretation, the names "Latte", "Norwegian espresso" and "Italian espresso" have been superimposed. The three segments are indicated by different symbols (colors), blue squares represent the latte segment, red crosses the Italian espresso and green pluses the Norwegian espresso.

<<Please, place here figure 4>>

4.3. *Investigation of consumer factors using PLS-DA*

Clusters identified by visual segmentation using PCA (2 clusters, almost identical to two clusters obtained by FCM here) on the non standardised regression coefficients were properly interpreted in Asioli, Almli, et al. (2016b). Focus will therefore be on the 3-clusters solution identified using standardised variables in the FCM. Table 3 describes the significant consumer factors for PLS-DA based on FCM model on standardized coefficients data.

Table 3 – Significant consumer factors for PLS-DA based on FCM model based on standardized coefficients data. In the column to the left are listed the significant consumer factors for the PLS-DA based on the FCM model. The column to the right gives a more detailed description of the consumer factors.

<<Please, place here table 3>>

The three segments presented in Figure 5 are referred to as "Latte Group" (right side), "Norwegian Espresso Group" (bottom left side) and "Italian Espresso Group" (top left side).

21

The explained variances for the first two components were 18% and 9% for X and 23% and 6% for Y.

Figure 5 – Correlation loadings with significant consumer factors from PLS-DA model based on three clusters identified with FCM based on standardized coefficients data. The factors/words that overlap and are therefore not easily readable in the left top part (i.e. "Italian espresso group") are: Coffee intensity, Friele IC and Home IC; in the left bottom part (i.e. "Norwegian espresso group") are Gender, Work/Un C, Regular C and Black: and, in the right part (i.e. "Latte group") are: Energy drink, Calories and Starbucks IC.

<<Please, place here figure 5>>

Segments characteristics

The three consumer groups (see Figure 5) differ for several consumer factors. Consumers belonging to the "Latte group" which correspond to the "Latte group" of Strategy one (see section 4.3.1) pay less attention to *calories* content compared to consumers belonging to the other groups.

In terms of warm coffee habits consumers belonging to "Latte group" prefer latte (i.e. Latte C) and mocca (i.e. Mocca C) while consumers belonging to "Norwegian espresso group" prefer to drink regular coffee (i.e. Regular C), black coffee (i.e. Black), warm coffee and drink warm coffee at work or university (i.e. Work/Un. C). Consumers belonging to the "Italian espresso group" have preference for espresso (i.e. Espresso C) and coffee intensity (i.e. Coffee intensity).

22

The characterization of the two Espresso groups is fully coherent with the location in the plot of the two segments. In addition, in terms of *iced coffee habits,* while consumers belonging to "Latte" group have a higher preference for energy products (i.e. Energy IC and Energy drink) and Starbucks IC products, consumers belonging to "Italian espresso group" prefer to drink product at home (i.e. Home IC) and Friele IC. Finally, in terms of *socio-demographics*, gender is significantly correlated to "Norwegian espresso group", meaning that males tend to belong to this group.

5.   DISCUSSION & CONCLUSIONS

This study compared two different *a posteriori* segmentation strategies for investigating individual differences among consumers. Data from a CE investigating consumers' preferences for iced coffee products in Norway were used. Strategy one is based on the so-called LCL model while Strategy two is based on analyzing and clustering the regression coefficients extracted using a ML model using either inspecting individuals using PCA plots or by the use of a more formal clustering procedure, for instance Fuzzy C Means (FCM). The most important difference between the two strategies is that Strategy one is based on assuming that all consumers can be split in subgroups with the same utility function for all, while Strategy two assumes a random regression vector making the utility different for each consumer.

Strategy one is a method which gives a suggested splitting into segments, regression coefficients for each segment and membership values to all segments. Strategy two on the other hand, provides regression coefficients for each individual that can be used for plotting and interpretation based on the PCA without any focus on segmentaton. This is useful in itself

23

for revealing the structure of the individual dfferences. In fact, segmentation can be based on the PCA plot directly using so-called interpretation based clustering (i.e. clustering based on which types of consumer preference patterns one is interested in studying). More details on interpretation based clustering can be found in Almli et al. (2011). Strategy two opens up for flexibility in terms of different weighting of the effects, which may open the possibility for additional interpretation and segmentation tools. This was for illustration here done by standardization of the main effects in the FCM approach, but other options may be used if other ways of weighing of the contributions is more natural from prior knowledge. The same possibility does not exist for standard LCL model (Strategy one) since it is invariant to the scale of the input variables. The Strategy two can therefore for these reasons be considered more flexible and also equipped with more tools for interpretation than Strategy one. A possible drawback with the Strategy two, is that ML assumes a continuous distribution of the regression vector, and if there are clear segments among the consumers without a continuum of preferences, this may not be a natural assumption.

In both Strategies one and two, external consumer factors can be related to the segments for improved interpretation. In such cases, the Strategy one becomes a two-stage process based on first finding the segments and then afterwards relating the segments to the consumer factors. Strategy two will on the other hand be a three step process: first the regression coefficients are estimated, then the coefficients are used for segmentation before the segments are related to the consumer factors. Sequential procedures have the advantage that each step can be interpreted and understood before the next step is taken. It should be mentiond that the scores from the PCA can also be related directly to consumer factors using standard PLS regression and without going via segments.

The final segmentation results for the two strategies gave for this data set more or less the same results, but the PCA plot of the regression coefficients provided additional insights also for understanding why Strategy one did not converge.

REFERENCES

Allenby, G. M., Arora, N., & Ginter, J. L. (1998). On the Heterogeneity of Demand. *Journal of Marketing Research*, *35*(3), 384–389.

Almli, V., Næs, T., Enderli, G., Sulmont-Rossé, C., Issanchou, S., & Hersleth, M. (2011). Consumers' acceptance of innovations in traditional cheese. A comparative study in France and Norway. *Appetite*, *57*(1), 110–120.

Asioli, D., Almli, V., & Næs, T. (2016). Comparison of two different strategies for investigating individual differences among consumers in choice experiments. A case study

based on preferences for iced coffee in Norway. *Food Quality and Preference*, *54*, 79–89.

Asioli, D., Næs, T., Granli, B. S., & Almli, V. (2014). Consumer preferences for iced coffee determined by conjoint analysis: an exploratory study with Norwegian consumers. *International Journal of Food Science & Technology*, *49*(6), 1565–1571.

Asioli, D., Næs, T., Øvrum, A., & Almli, V. (2016). Comparison of rating-based and choice-based conjoint analysis models. A case study based on preferences for iced coffee in Norway. *Food Quality and Preference*, *48*, 174–184.

Bech, M., & Gyrd-Hansen, D. (2005). Effects coding in discrete choice experiments. *Health Economics*, *14*(10), 1079–1083.

Berget, I., Mevik, B.-H., & Næs, T. (2008). New modifications and applications of fuzzy - means methodology. *Computational Statistics & Data Analysis*, *52*(5), 2403–2418.

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms:* Plenum Press.

Bhat, C. R. (1997). An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. *Transportation Science*, *31*(1), 34–48.

Caputo, V., Nayga, R. M., & Scarpa, R. (2013). Food miles or carbon emissions? Exploring labelling preference for food transport footprint with a stated choice study. *Australian Journal of Agricultural and Resource Economics*, *57*(4), 465–482. http://doi.org/10.1111/1467-8489.12014

Carson, R., Louviere, J., Anderson, D., Arabie, P., Bunch, D., Hensher, D., … Wiley, J. (1994). Experimental analysis of choice. *Marketing Letters*, *5*(4), 351–367.

Combris, P., Bazoche, P., Giraud-Héraud, E., & Issanchou, S. (2009). Food choices: What do we learn from combining sensory and economic experiments? *Food Quality and Preference*, *20*(8), 550–557.

Endrizzi, I., Gasperi, F., Rødbotten, M., & Næs, T. (2014). Interpretation, validation and segmentation of preference mapping models. *Food Quality and Preference*, *32*, 198–209.

Frewer, L. J., & van Trijp, H. (2007). *Understanding Consumers of Food Products*. Cambridge, UK: Woodhead Publishing, CRC Press.

Greene, W., & Hensher, D. (2003). A latent class model for discrete choice analysis: contrasts with mixed logitNo Title. *Transportation Research Part B: Methodological*, *37*(8), 681–698.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, *17*(2), 107–145.

Hess, S., Ben-Akiva, M., Gopinath, D., & Walker, J. (2011). *Advantages of latent class over continuous mixture of Logit model*.

Louviere, Hensher, & Swait. (2000). *Stated choice methods. Analysis and applications*. Cambridge: Cambridge University Press.

MacFie, H. (2007). *Consumer-Led Food Product Development*. (MacFie H, Ed.). Cambridge: Woodhead Publishing Limitated.

Martens, H., & Martens, M. (2000). Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Quality and Preference*, *11*(1–2), 5–16.

Næs, T., Kubberød, E., & Sivertsen, H. (2001). Identifying and interpreting market segments

using conjoint analysis. *Food Quality and Preference*, *12*(2), 133–143.
http://doi.org/http://dx.doi.org/10.1016/S0950-3293(00)00039-2

Næs, Brockhoff, P., & Tomic, O. (2010). *Statistics for sensory and consumer science*. (Wiley,
Ed.). Chichester, UK.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes
in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study.
Structural Equation Modeling: A Multidisciplinary Journal, *14*(4), 535–569.

Ortega, D. L., Wang, H. H., Wu, L., & Olynk, N. J. (2011). Modeling heterogeneity in
consumer preferences for select food safety attributes in China. *Food Policy*, *36*(2), 318–324.

Pacifico, D., & Hong il, Y. (2013). lclogit: A Stata command for fitting latent-class
conditional logit models via the expectation-maximization algorithm. *The Stata Journal*,
13(3), 625–639.

Pacifico, D., & Yoo, H. Il. (2012). *clogit: a Stata module for estimating a mixed logit model
with discrete mixing distribution via the Expectation-Maximization algorithm.*

Peschel, A. O., Grebitus, C., Steiner, B., & Veeman, M. (2016). How does consumer
knowledge affect environmentally sustainable choices? Evidence from a cross-country latent
class analysis of food labels. *Appetite*, *106*, 78–91.

Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode
choice data. *Applied Economics*, *41*(22), 2915–2924.

Smith, W. R. (1956). Product Differentiation and Market Segmentation as Alternative
Marketing Strategies. *Journal of Marketing*, *21*(1), 3–8.

Ståhle, L., & Wold, S. (1987). Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *Journal of Chemometrics*, *1*(3), 185–196.

Train. (2009). *Discrete choice methods with simulation*. (C. U. Press, Ed.), *New York* (Vol. 47). Cambridge University Press.

Train, K. E. (2008). {EM} Algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling*, *1*(1), 40–69.

Wajrock, S., Antille, N., Rytz, A., Pineau, N., & Hager, C. (2008). Partitioning methods outperform hierarchical methods for clustering consumers in preference mapping. *Food Quality and Preference*, *19*(7), 662–669.

Wedel, M., & Kamakura, W. (2002). Introduction to the Special Issue on Market Segmentation. *International Journal of Research in Marketing*, *19*(3), 181–183.

Highlights

- Two different clustering methods using choice data are compared.

- We tested the strategies on different product profiles of iced coffees with Norwegian consumers.

- Strategy one is based on the classical Latent Class Logit (LCL) model.

- Strategy two uses Mixed Logit (ML) model combined with PCA for visual segmentation or with automatic clustering.

- Strategy two is preferred for its flexibility and equipped with more tools for interpretation.

Table 1 – Conjoint factors, levels and codes.

FACTOR	LEVELS	CODES
Coffee	Latte	-1
	Espresso	1
Calories	60 kcal/100 ml	-1
	90 kcal/100 ml	1
Origin	Norway	-1
	Italy	1
Price	17 NOK	-1
	23 NOK	0
	29 NOK	1

Table 2 - Estimated parameters for LCL and ML models with conjoint variables' main effects and interactions. The two columns to the left correspond to the estimated regression coefficients for the two cluster based on LCL model while the two columns to the right refer to the population effects based on the ML model (mean and SD).

FACTOR	LATENT CLASS LOGIT MODEL		MIXED LOGIT MODEL	
	Group 1 average Estimate	Group 2 average Estimate	Group average Estimate	Individual variation (SD)
Coffee	0.916***	-0.980**	-0.183	1.881***
Calories	-0.170	-0.500***	-0.571***	0.557***
Origin	-0.159	-0.548***	-0.281**	0.666***
Price	-0.160***	-0.098***	-1.06***	0.596***
Coffee*Calories	-0.066	0.261	0.061	0.204
Coffee*Origin	0.151	0.330	0.162	0.306
Coffee*Price	0.062**	0.118***	0.229*	0.007
Calories*Origin	0.060	-0.191	0.046	0.042
Calories*Price	-0.018	0.049	-0.062	0.073
Origin*Price	-0.009	-0.037	-0.111	0.052

*, ** and *** indicate significant effects at 0.05, 0.01 and 0.001 levels, respectively.

Number of choice observations: 2448

Number of consumers: 102

Table 3 – Significant consumer factors for PLS-DA based on FCM model based on standardized coefficients data. In the column to the left are listed the significant consumer factors for the PLS-DA based on the FCM model. The column to the right gives a more detailed description of the consumer factors.

PLS-DA based on FCM model, 3 clusters	Consumer factor description
Latte C	Preference for latte warm coffee
Calories	Preference for iced coffee products with high calories content
Coffee intensity	Preference for coffee intensity warm coffee
Espresso C	Preference for espresso warm coffee
Regular C	Preference for regular warm coffee
Black	Preference for black warm coffee
Friele IC	Preference for iced coffee with Friele brand
Home IC	Preference for iced coffee consumed at home
Work/Un. C	Preference for warm coffee consumed at work or university
Gender	Males' preference for some coffee types
Starbucks IC	Preference for iced coffee with Starbucks brand
Energy drink	Preference for drinks with high energy content
Mocca C	Preference for mocca warm coffee
Warm coffee	Preference for warm coffee

Figure 1 – PCA correlation loadings plot for PC-1 and PC-2 on individual ML parameter estimates from choice data. The interactions effect that overlap (i.e. non clearly understandable) in the middle of the plot are: espresso_highcalories, espresso_italy, espresso_highprice, highcalories_italy, highcalories_highprice and italy_highprice.
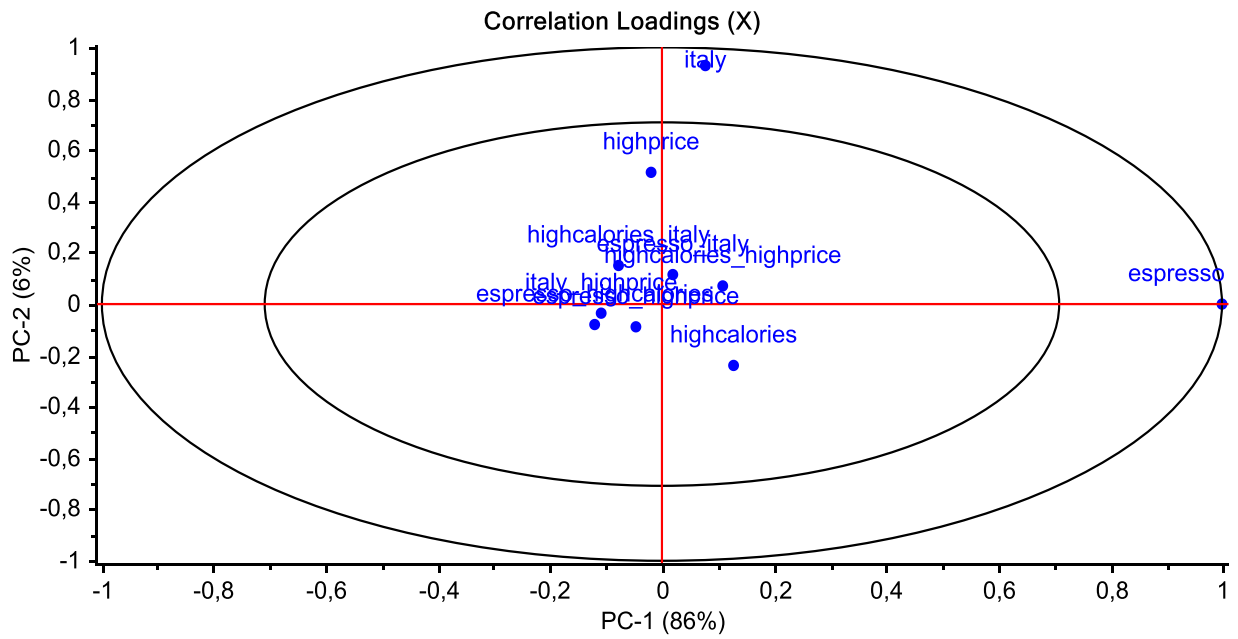
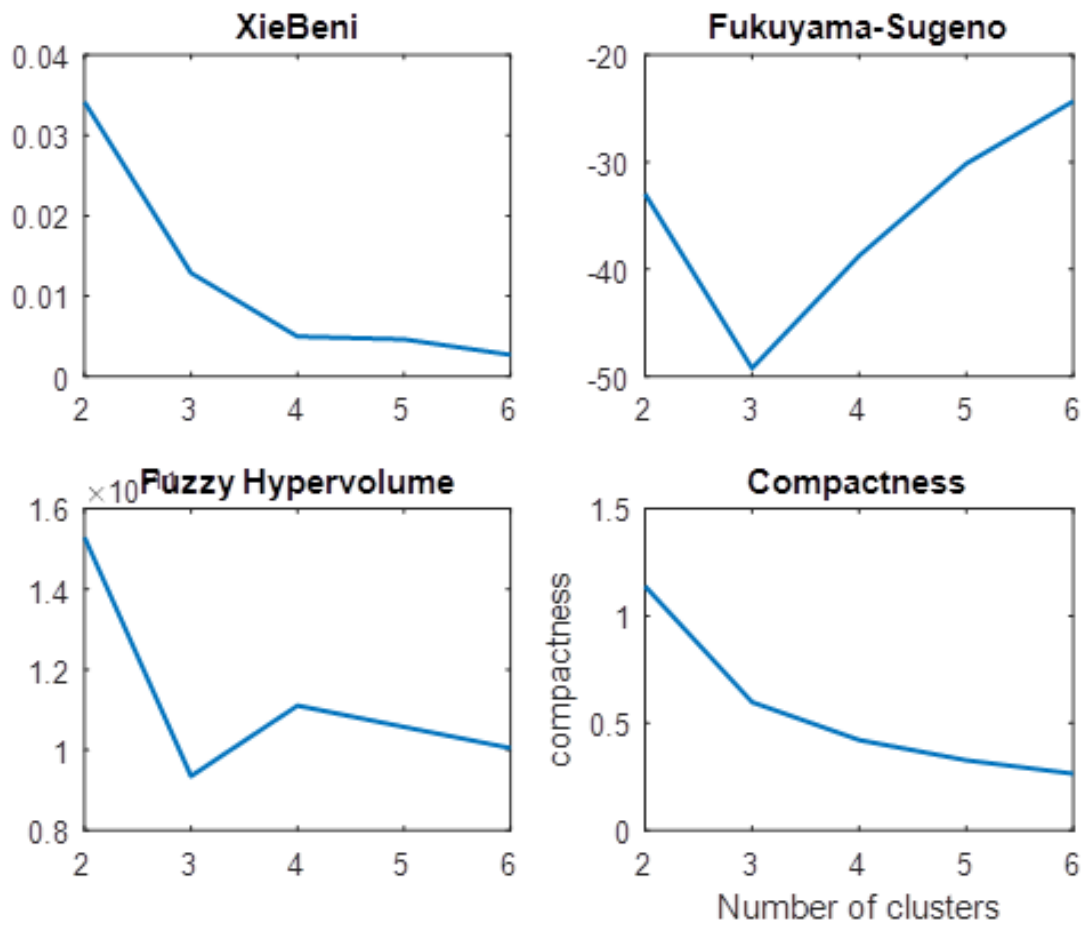**Figure 2 - Plots of the four indices tested for cluster number in FCM (non-standardised data)**

Figure 3 - The results from FCM for three clusters plotted in the PCA scores plot. To ease interpretation, the names "Latte", "Espresso" and "Segment 3" have been superimposed. The three segments are indicated by different symbols (colors), blue squares represent the latte segment, green triangles the espresso segment and red circles the third, middle segment.
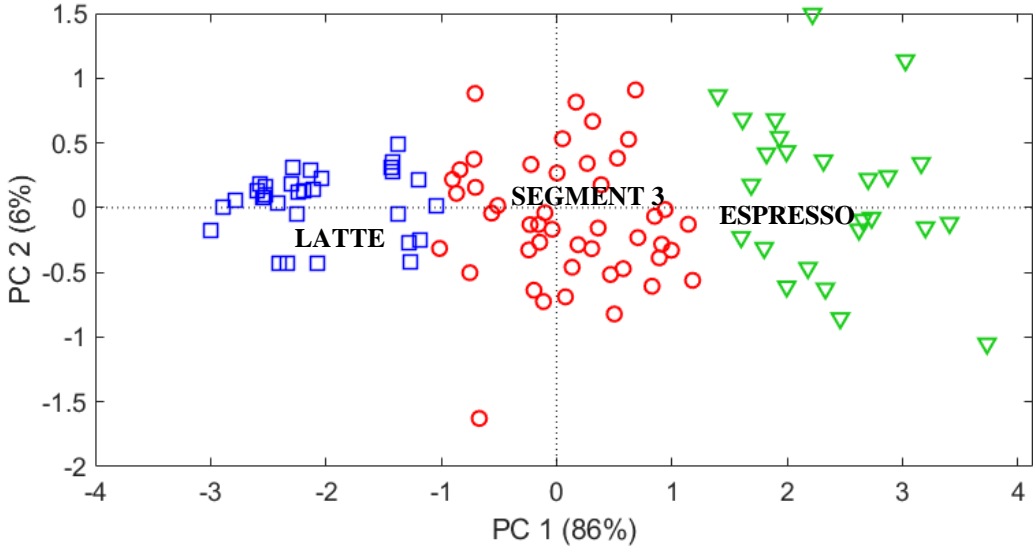
Figure 4 – Score plot obtained with unstandardized data, and clusters obtained with FCM on standardized data and S=3. To ease interpretation, the names "Latte", "Norwegian espresso" and "Italian espresso" have been superimposed. The three segments are indicated by different symbols (colors), blue squares represent the latte segment, red crosses the Italian espresso and green pluses the Norwegian espresso.
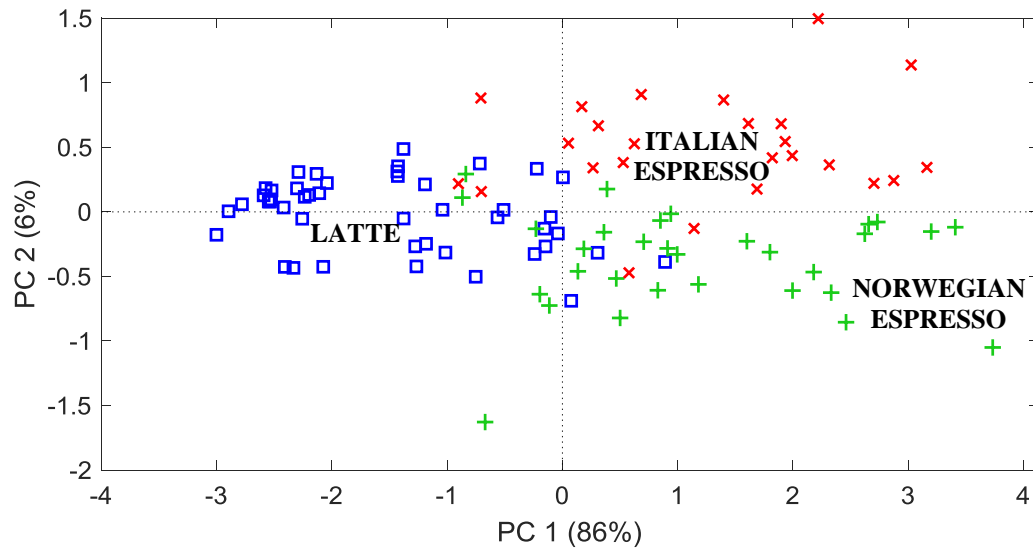
Figure 5 – Correlation loadings with significant consumer factors from PLS-DA model based on three clusters identified with FCM based on standardized coefficients data. The factors/words that overlap and are therefore not easily readable in the left top part (i.e. "Italian espresso group") are: Coffee intensity, Friele IC and Home IC; in the left bottom part (i.e. "Norwegian espresso group") are Gender, Work/Un C, Regular C and Black: and, in the right part (i.e. "Latte group") are: Energy drink, Calories and Starbucks IC.