Choice models with mixtures: an application to a cocktail experiment

Peter Goos^{1,2} and Hajar Hamidouche¹

¹Department of Biosystems, KU Leuven, Belgium

²Department of Engineering Management, University of Antwerp, Belgium

April 12, 2019

Abstract

Data from choice experiments are analyzed using choice models. Proper modeling allows a good understanding of the consumer preferences, the identification of consumer segments and the optimization of products, and, eventually, offers opportunities for competitive advantage. However, in the analysis of choice data, products that are mixtures of ingredients have been largely overlooked. In this paper, we combine traditional choice models and traditional mixture models. We apply the resulting mixture choice models to data from a real-life experiment in which consumers made pair-wise comparisons involving seven cocktails. For the choice models, we first assume consumer homogeneity. Next, we allow for heterogeneity among individuals. Therefore, we discuss the multinomial logit model, the mixed logit model and the latent class model. For identifying segments, besides the latent class model, we explore various two-stage segmentation approaches in which forces, hierarchical Bayes estimates and Firth individual-level estimates are used as input for a cluster analysis. The results show that the mixed logit model and the latent class model describe the data better than the multinomial logit model. The optimal cocktails for the segments we identified using the various methods are not identical, but show many similarities.

Keywords: cluster analysis, Firth estimates, forces, hierarchical Bayes estimates, latent class model, mixed logit model, mixtures, multinomial logit model, Scheffé model

1 Introduction

Choice experiments are frequently used to perform systematic studies of consumer preferences. In these experiments, the properties of products are systematically varied according to an experimental design. After a choice experiment has been conducted, the data are analyzed using choice models. A good understanding of consumer preferences offers opportunities for competitive advantage as it enables the identification of consumer segments and the optimization of products or services.

The literature about discrete choice models is very extensive (Hensher and Johnson, 2018; Louviere and Woodworth, 1983; McFadden, 1986; Ben-Akiva and Boccara, 1995; McFadden and Train, 2000; Louviere et al., 2000; Hensher et al., 2005) and applications can be found in marketing (Kamakura and Russell, 1989), environmental economics (Bennett and Blamey, 2001), transport (Kupfer et al., 2016), health care (Lancsar and Louviere, 2008; Luyten et al., 2015), ecology (Vardakis et al., 2015) and many other disciplines. However, in the analysis of choice data, products that are mixtures of ingredients have been largely overlooked. This is surprising as a great share of products, such as shampoos, potato chips, cakes and cocktails, actually involve mixtures.

The literature on the modeling of data from traditional mixture experiments is also substantial. Scheffé’s pioneering article “*Experiments with mixtures*” from 1958 was the impetus for the interest in the design and analysis of mixture experiments. A chronological list of literature on mixture experiments from 1953 to 2002 can be found in Cornell (2011). Cornell (2011) applies models for data from mixture experiments to fruit punch, chick feeding, sports drinks, paper coatings, and many other products. Many of these products are food products.

In this paper, we combine commonly used choice models and traditional mixture models. We apply the resulting models to data from a real-life experiment in which consumers made pair-wise comparisons involving seven cocktails. We first discuss the multinomial logit model, which is suitable when individuals are homogeneous. Next, to deal with scenarios in which individuals have heterogeneous preferences, we discuss the mixed logit model and the latent class model. Also, we study two-stage customer segmentation approaches in which forces, hierarchical Bayes estimates and Firth individual-level estimates, and a combination of these, are used as input for a cluster analysis.

The rest of the paper is organized as follows. In Section 2, we briefly discuss discrete choice modeling, two-stage customer segmentation and traditional linear models for data from mixture experiments. In Section 3, we combine traditional choice models and mixture models. Sections 4 and 5 apply the new mixture-choice models to a real-life cocktail experiment. In Section 4, we focus on the most important response, taste preference, while, in

Section 5, we study the color preference. Finally, we draw conclusions and provide suggestions for future research in Section 6.

2 Discrete choice modeling and mixtures

2.1 Discrete choice modeling

Several models have been built to describe which factors affect the choices made by respondents in choice experiments or stated choice studies. Train (2009) provides a review of the most commonly used models for choice behavior. All these models build on the utility-maximization theory, which states that, when an individual is confronted with several options, the chosen option is the one which maximizes the perceived utility. A collection of alternatives between which a consumer has to choose is generally referred to as the choice set. The utility perceived by a respondent n of an alternative i is generally denoted by U_{ni} and decomposed in a systematic part V_{ni} and a random part ϵ_{ni} :

$$U_{ni} = V_{ni} + \epsilon_{ni}, \quad (1)$$

where n refers to the individual and i refers to the alternative (Train, 2009). The systematic part V_{ni} is a function of observable factors and the effects of these factors. In the event all individuals have the same preferences, the effects of the factors can be represented by a single parameter vector β . In the event all individuals have different preferences, every individual has his/her own parameter vector β_n . The most important observable factors are attributes of the products under investigation, but they can include characteristics of the respondents, such as their age and gender, as well.

According to the utility-maximization theory, the probability P_{nij} that alternative i will be chosen by respondent n from a pair of two alternatives, i and j , equals the probability that the perceived utility for alternative i is higher than the perceived utility for alternative j (Train, 2009):

$$P_{nij} = P(U_{ni} > U_{nj}) = P(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj}) = P(V_{ni} - V_{nj} > \epsilon_{nj} - \epsilon_{ni}). \quad (2)$$

The last expression in this equation shows that only differences between utilities are relevant. The probability P_{ni} that alternative i will be chosen by respondent n from a set of J alternatives, $1, \dots, J$, equals the probability that the perceived utility for alternative i is higher than the perceived utilities of all other alternatives. That probability can be written

as

$$P_{ni} = P(U_{ni} > \max(U_{n1}, \dots, U_{n,i-1}, U_{n,i+1}, \dots, U_{nJ})). \quad (3)$$

Equations (2) and (3) form the basis for the derivation of the multinomial logit model, the mixed logit model and the latent class model. These commonly used models differ in their assumptions about the preference homogeneity/heterogeneity of the individuals.

2.1.1 Multinomial logit model

The simplest model for choice data is the multinomial logit model (MNL). This model is built by assuming that the random components of the perceived utilities are independently and identically distributed according to an extreme value type 1 distribution, also known as the Gumbel distribution. In addition, when applying the MNL to a set of data coming from multiple respondents, these respondents are assumed to have homogeneous preferences, meaning that every product attribute affects every respondent's choices in the same way. As a result, the parameter vector β describing the effects of the factors on the perceived utility is the same for each individual n . The probability that alternative i is chosen in a choice set containing J alternatives can then be written in closed form as

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}}, \quad (4)$$

where

$$V_{ni} = f(x_i, \beta),$$

and x_i represents the vector of alternative i 's factor values. Traditionally, maximum likelihood estimation has been used for estimating the parameter vector β . However, instead of maximum likelihood estimates, Kessels et al. (2019) recommend the use of Firth penalized-likelihood estimates (Firth, 1993) for both aggregate and individual data, for two reasons. First, if the data shows separation, which is quite common in small samples involving binary responses, estimates based on maximum likelihood do not exist. Second, even if the maximum likelihood estimates exist, they tend to be substantially biased when the sample is small. The penalized-likelihood method of Firth overcomes these two problems. For this reason, in this paper, we use the Firth estimation approach for the multinomial logit model. Firth's approach is available for generalized linear models, such as logistic regression models and Poisson regression models, in multiple software packages nowadays.

Generally, the multinomial logit model is fitted to the aggregate data set, which combines the data collected from all available respondents. If there is sufficient data for each individual respondent, the MNL can however also be applied to the choice data from any

individual separately. This produces individual-level estimates of the parameters in β which quantify the preferences of an individual respondent, and therefore offers the possibility to study whether the preferences of the respondents are homogeneous or heterogeneous. More specifically, the individual-level estimates can be used to cluster the respondents to attempt to find sets of respondents with similar preferences. This procedure is generally referred to as segmentation and is discussed in more detail in Section 2.2.

In any case, the MNLM's assumption that preferences of individuals are homogeneous is not always realistic, as was shown by Greene and Hensher (2003) and Shen (2009) who modeled the choice between road types and transportation modes, respectively. This has led to the development and widespread use of the mixed logit model and the latent class model.

2.1.2 Mixed logit model

Unlike the MNLM, the mixed logit model (MLM) allows for preference heterogeneity between individuals. More specifically, the MLM assumes that each individual n has his/her own parameter vector β_n , so that

$$V_{ni} = f(x_i, \beta_n). \quad (5)$$

The choice probability in Equation (4) is then conditional on β_n and valid only for respondent n . In the MLM, the respondents are assumed to be randomly selected from a population, which is formalized by stating that the respondents' individual parameter vectors β_n are drawn from a continuous distribution $\pi(\beta|\phi)$. Oftentimes, $\pi(\beta|\phi)$ is the normal or lognormal distribution. The unconditional probability that a certain product i is chosen from a set with J alternatives is then the weighted average of the conditional probabilities obtained from Equations (4) and (5). That weighted average is calculated using an integral (Train, 2009):

$$P_{ni} = \int \frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}} \pi(\beta|\phi) d\beta. \quad (6)$$

In this paper, we used the normal distribution to describe how heterogeneous the respondents' preferences can be. This makes sense when there is no prior knowledge concerning the signs of the parameters in the vector β_n . The vector ϕ contains the so-called population-level parameters and describes the properties of the assumed normal distribution. It involves at least the mean vector β for all individual β_n vectors and a vector of standard deviations σ_β . It is possible to allow for dependence between the parameters within the β_n vectors, in which case the vector ϕ also involves covariances. In this paper, for lack of *a priori* knowledge, we treat the elements of the β_n vectors as independent.

There exists no closed-form solution for the unconditional choice probability in Equ-

tion (6), which implies that simulations are needed to estimate the model. The two most commonly used techniques to estimate the MLM are hierarchical Bayes estimation and maximum simulated likelihood (Train, 2001). In this paper, we use hierarchical Bayes estimation.

Estimating a MLM results in two pieces of output. The first piece of output reports estimates of the population-level parameters contained within the vector ϕ . In the event $\pi(\beta|\phi)$ is the normal distribution, the estimates provide information concerning the average of all respondents' β_n vectors and the standard deviations. The second piece of output includes estimates of the respondents' individual β_n vectors. These form a second type of individual-level estimates and can also be used to cluster the respondents to attempt to find sets of respondents with similar preferences.

While applying the MNLM to individual respondents' data only works if individual respondents have evaluated at least as many choice sets as there are parameters in the model, this is not the case for the MLM. When the goal is to obtain individual-level estimates, the MLM thus requires fewer data points per respondent than the MNLM. This is certainly an attractive feature, but it is important to realize that this possibility is due to the assumption that the respondents all come from the same population. So, **the MLM assumes** that the respondents have different preferences, but, at the same time, it involves a population distribution $\pi(\beta|\phi)$ that restricts the extent to which the preferences can differ from one respondent to the other. When applying the MNLM to individual respondents' data, there is no such restriction.

2.1.3 Latent class model

The MNLM and the MLM can be viewed as two ends of a spectrum. When applied to a full set of data from multiple respondents, the MNLM assumes that all respondents have the same preferences, which can be summarized using a single vector β . In contrast, the MLM assumes that every respondent is different. This is reflected by the fact that every respondent has his/her own parameter vector β_n . The latent class model can be viewed as a middle course between the MNLM and the MLM.

In the latent class model (LCM), it is assumed that a limited number of segments or classes are sufficient to account for the preference heterogeneity among the respondents. The respondents belonging to a given segment or class are assumed to have the same preferences. For each segment or class c , the latent class model therefore involves a parameter vector β_c which describes the preferences of all respondents who belong to that segment. The latent class probability that an individual n belonging to class c chooses alternative i from a choice

set comprising J alternatives is

$$P_{ni|c} = \frac{e^{V_{ni|c}}}{\sum_{j=1}^J e^{V_{nj|c}}}, \quad (7)$$

where

$$V_{ni|c} = f(x_i, \beta_c). \quad (8)$$

This expression, which is conditional on β_c , corresponds to the multinomial logit probability from Equation (4) applied to the respondents within class c . The unconditional probability that an individual n chooses alternative i can be calculated as

$$P_{ni} = \sum_{c=1}^C P_{ni|c} P_{nc}, \quad (9)$$

where

$$P_{nc} = \frac{e^{\theta_c z_n}}{\sum_{c=1}^C e^{\theta_c z_n}} \quad (10)$$

is the probability that respondent n belongs to class c , C is the number of classes considered, θ_c represents a vector of class-membership parameters and z_n is an optional set of characteristics related to individual n (Greene and Hensher, 2003). **In the event no individual characteristics are available, $\theta_c z_n$ is just a constant, the class-membership probability P_{nc} is independent of the respondents' properties, and we can drop the index n from that probability and denote it by P_c . That probability, P_c , can be interpreted as the size or the importance of segment c .**

One difficulty when using the latent class model is that the number of classes, C , must be determined. To select the most appropriate number of classes C , Nylund et al. (2007) recommend using the Bayesian information criterion (BIC). The optimal number of classes to use is the one that minimizes the BIC value.

2.2 One- and two-stage segmentation

To optimize products so that they are appealing to many consumers, it is important to identify large segments of consumers with the same preferences. For each of these consumer segments, a product variant can be developed. The ideal product variant for a segment c maximizes the systematic utility $V_{ni|c} = f(x_i, \beta_c)$, where β_c is the vector that quantifies the preferences of the consumers in the segment. In order to be able to adopt this strategy, the various consumer segments have to be identified.

The LCM identifies segments automatically. In addition, that model involves parameter vectors β_c which summarize the preferences of the respondents in each of the segments c . The

estimates for the vectors β_c can then be used to find the product settings x_i that optimize the systematic utility $V_{ni|c}$. Therefore, the LCM is a one-stage method for segmentation and product differentiation.

As the LCM requires more sophisticated software than two-stage methods, Crabbe et al. (2013) analyzed the performance of ten two-stage segmentation methods (Crabbe et al., 2013). A two-stage segmentation method starts with the calculation of indicator values for each respondent in the choice experiment. Next, the respondents are classified into clusters, segments or classes by means of a cluster analysis or a latent class model. Crabbe et al. (2013) show that hierarchical Bayes individual-level estimates as well as so-called forces as input for a cluster analysis perform well in terms of success rates, correctly classified individuals and mean estimation error. Kessels et al. (2019) demonstrate that Firth individual-level estimates, obtained by fitting the MNLM to data from individual respondents using Firth’s penalized-likelihood method, also allow for a reliable segmentation of the respondents.

Hierarchical Bayes (HB) individual-level estimates are a by-product of the hierarchical Bayes estimation procedure for the MLM applied to the complete data set for all respondents. Individuals with approximately the same HB individual-level estimates have similar preferences and therefore can be assumed to belong to the same segment. Kessels et al. (2019) compared the Firth individual-level estimates with HB individual-level estimates in the context of consumer segmentation, and **find that segment recovery based on Firth individual-level estimates is more precise overall**. Therefore, in addition to HB individual-level estimates, we also use the Firth individual-level estimates as indicator values for a cluster analysis.

The third kind of indicator values we utilize as input for a cluster analysis are the so-called forces. In simple words, the forces quantify the influence an individual respondent exerts on the estimates of the parameters in the aggregate MNLM, i.e., the MNLM fit to the complete set of data from all respondents. The formula used for the calculation of the forces of an individual n is

$$\bar{\Delta}_n = \sum_{j=1}^J \sum_{s=1}^S \frac{\Delta_{njs}}{J \times S}, \quad (11)$$

where J is the number of alternatives in a choice set, S is the number of choice sets evaluated by an individual respondent and Δ_{njs} represents the force applied by alternative j evaluated by individual n in choice set s . For more technical details on the calculation of forces, we refer to Crabbe et al. (2013). The idea is that individuals who exert the same kind of influence on the MNLM’s parameter estimates are alike and therefore belong to the same segment.

Finally, we also investigate whether there is added value in combining the three sets of indicator values discussed above (HB individual-level estimates, Firth individual-level

estimates and forces) into one extensive set of indicator values. A cluster analysis of this combined set of indicator values uses all available individual-level information to identify segments, and therefore has the potential to lead to the most trustworthy segmentation.

For our application in Section 4, we use Ward’s hierarchical clustering method. As any hierarchical clustering method, Ward’s method has the advantage that the number of clusters or segments does not have to be specified in advance. Moreover, Punj and Stewart (1983) found that Ward’s method, in which clusters are merged by minimizing the total within-group or within-cluster sum of squares, outperforms the other hierarchical approaches for most data. When N respondents have to be clustered, hierarchical clustering starts from N clusters (i.e., each initial cluster comprises a single respondent). The clustering procedure then sequentially merges the two most similar clusters, until a user-specified stopping rule is met. We used the distance graph to decide on when to stop the clustering. In this graph, the distance between the joined clusters is depicted on the vertical axis and the number of clusters appears on the horizontal axis. We stopped the clustering when the distance between the joined clusters shows a sudden increase. This approach is akin to the use of the scree plot in principal component analysis to determine a suitable number of principal components.

2.3 Mixtures

Many products and services are mixtures of two or more ingredients. A few examples are cakes, shampoos and cocktails. In traditional mixture experiments, the output is assumed to depend only on the proportions of the ingredients in the mixture and not on the amount of the mixture (Cornell, 2011; Goos and Jones, 2011). The restrictions inherent to the mixture problem have led to the well-known Scheffé models. The first restriction states that the proportions $x_1, \dots, x_q$ of the q ingredients in the mixture should sum up to 1. Moreover, the proportions should be positive. These restrictions can be written as

$$\sum_{i=1}^q x_i = 1, \tag{12}$$

and

$$x_i \geq 0, \quad i = 1, \dots, q, \tag{13}$$

respectively. As a result of these constraints, the experimental region for mixture experiments is a simplex.

The first-order Scheffé model for a continuous response Y is

$$Y = \sum_{i=1}^q \beta_i x_i + \epsilon, \quad (14)$$

whereas the second-order Scheffé model is given by

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \epsilon. \quad (15)$$

Finally, the most commonly used third-order Scheffé model, also known as the special-cubic model, is

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \sum_{i=1}^{q-2} \sum_{j=i+1}^{q-1} \sum_{k=j+1}^q \beta_{ijk} x_i x_j x_k + \epsilon. \quad (16)$$

Details about the derivation of these models can be found in Scheffé (1958) and Cornell (2011). In the Scheffé models, the coefficient β_i can be interpreted as the expected response when the proportion of ingredient i equals 100%. If interactions between ingredients are expected, the second-order or special-cubic Scheffé model are appropriate. The parameter β_{ij} represents the interaction effect between ingredients i and j . The ingredients i and j are said to be in synergism/antagonism if the expected response of a mixture is greater/smaller than the weighted sum of β_i and β_j . In the case of synergism, the parameter β_{ij} has a value greater than 0. In the case of antagonism, it has a negative value.

Experimental designs suitable for estimating the above first-order, second-order and special-cubic models have been proposed by Scheffé (1958, 1963). These experimental designs are called simplex-lattice and simplex-centroid designs. Some of the designs have been proven to be D-optimal for the Scheffé models in Equations (14)-(16) (Kiefer, 1961a,b; Atwood, 1969), implying that the designs allow a precise model estimation. A review of the optimality properties of these traditional designs for mixture experiments can be found in Goos et al. (2016).

An example of a simplex-lattice design is shown in Figure 1a. That design is the {3,2} simplex-lattice design, suitable for experiments with three ingredients and D-optimal for estimating the second-order Scheffé model. The simplex-centroid design for a mixture experiment involving three ingredients is shown in Figure 1b. That design is D-optimal for estimating the three-ingredient special-cubic model.

In many mixture experiments, there are lower bounds on the proportions of the ingredients. In that case, the experimental region is a smaller simplex, embedded within the original one defined by Equations (12) and (13). The simplex-lattice and simplex-centroid designs

(a) A $\{3,2\}$ simplex-lattice design.

(b) A $\{3,2\}$ simplex-centroid design.

Figure 1: Two traditional experimental designs for mixture experiments with three ingredients.

are then shrunk to fit within the smaller simplex, because the resulting shrunk designs retain their optimality properties.

3 Choice models for mixtures

To model choices between mixtures that differ in terms of their ingredients' proportions, we need to embed the Scheffé models from Section 2.3 in the choice models discussed in Section 2.1. In other words, the Scheffé models should replace the systematic parts $f(x_i, \beta)$, $f(x_i, \beta_n)$ and $f(x_i, \beta_c)$ of the MNLM, the MLM and the LCM, respectively.

There is, however, one technical difficulty, due to the constraint that all ingredient proportions sum to one (Equation (12)). To see this, suppose that we model the systematic utility of mixture i evaluated by individual n using the first-order Scheffé model, assuming that all respondents have the same preference parameters, $\beta_1, \dots, \beta_q$, for the q ingredients:

$$V_{ni} = f(x_i, \beta) = \sum_{k=1}^q \beta_k x_{ki}, \quad (17)$$

with x_{ki} the proportion of ingredient k in mixture i . Due to Equation (12), we have that $x_{qi} = 1 - x_{1i} - \dots - x_{(q-1)i}$, so that we can rewrite the model in Equation (17) as

$$\begin{aligned} V_{ni} &= \sum_{k=1}^{q-1} \beta_k x_{ki} + \beta_q (1 - x_{1i} - \dots - x_{(q-1)i}), \\ &= \beta_q + \sum_{k=1}^{q-1} (\beta_k - \beta_q) x_{ki}. \end{aligned} \quad (18)$$

This expression for the systematic utility of mixture i contains a constant term. However, as pointed out in Section 2.1, when it comes to making choices and maximizing utilities, it is differences in utility that matter, rather than absolute utilities. Consequently, any choice model in which the systematic utility involves a constant term is not identified and therefore inestimable. For this reason, also any choice model using a utility function involving the linear terms in the q ingredients of a mixture, such as the model in Equation (17), is inestimable.

This problem can be solved easily, by dropping one of the q linear terms in the ingredients. Which linear term is dropped is arbitrary: it does not affect the goodness of fit, nor any qualitative conclusion from the analysis. However, the coefficients of the $q - 1$ remaining linear terms have to be interpreted as effects relative to the effect of the omitted ingredient. In this article, we consistently drop the q th ingredient in our choice models for mixtures. The technique of dropping one of the linear terms in the q ingredients is analogous to the way in which classical regression analyses deal with the dummy variables corresponding to a categorical explanatory variable: such an analysis always uses one fewer dummy variable than there are levels for the categorical variables.

Incorporating the first-order, second-order and special-cubic Scheffé models in the MNLM, which assumes that all respondents have the same preferences, thus necessitates the use of the following expressions for the systematic utility:

$$V_{ni} = \sum_{k=1}^{q-1} (\beta_k - \beta_q) x_{ki}, \quad (19)$$

$$V_{ni} = \sum_{k=1}^{q-1} (\beta_k - \beta_q) x_{ki} + \sum_{k=1}^{q-1} \sum_{l=k+1}^q \beta_{kl} x_{ki} x_{li}, \quad (20)$$

and

$$V_{ni} = \sum_{k=1}^{q-1} (\beta_k - \beta_q) x_{ki} + \sum_{k=1}^{q-1} \sum_{l=k+1}^q \beta_{kl} x_{ki} x_{li} + \sum_{k=1}^{q-2} \sum_{l=k+1}^{q-1} \sum_{m=l+1}^q \beta_{klm} x_{ki} x_{li} x_{mi}. \quad (21)$$

In the event we want to estimate a MLM, we can use each of these three expressions too, but we have to add an extra index n to each model parameter. More specifically, we have to use β_{kn} , β_{qn} , β_{kln} and β_{klmn} instead of β_k , β_q , β_{kl} and β_{klm} , respectively, to allow the preferences to differ from respondent to respondent. For the LCM, the extra index n , corresponding to respondent n , has to be replaced by the index c corresponding to latent class or segment c . This allows the preferences to differ from segment to segment (but not from respondent to respondent).

4 Application to a cocktail experiment

We use the data of Courcoux and Séménou (1997) to illustrate the application of the mixture choice models introduced in the previous sections. We compare the various models, discuss the segments we identified and the mixtures that optimize the predicted utilities for the different segments.

4.1 The data

The data we use for our illustration come from an experiment involving seven cocktails containing mango juice, cassis and lime described by Courcoux and Séménou (1997). The seven cocktails studied are defined by a simplex-centroid design on a simplex-shaped experimental region. As there are lower bounds for the proportions of the three ingredients, the simplex-shaped experimental region is a shrunk one. The lower bounds for the proportions of mango, cassis and lime in the study equal 0.3, 0.15 and 0.1, respectively. As a result, the proportion of mango should be in the interval $[0.3, 0.75]$, while the proportion of cassis should be in the interval $[0.15, 0.6]$ and that of lime should be in the interval $[0.1, 0.55]$. So, all ingredient proportions have a range equal to 0.45. The experimental region and the seven cocktails are shown in Figure 2. In the figure, due to the lower bounds, the corners of the ternary plot do not represent three single-component mixtures, but the three ternary mixtures with the maximum proportions of either mango, cassis or lime. For example, cocktail 1 contains 75% of mango, 15% of cassis and 10% of lime. Cocktail 7, which corresponds to the centroid of the experimental region, involves 45% of mango, 30% of cassis and 25% of lime. **Using the seven cocktails in Figure 2 makes sense since the seven-point simplex-centroid design is the D- and I-optimal experimental design for the three-ingredient special-cubic Scheffé model (Goos et al., 2016).**

The study involved sixty respondents, thirty males and thirty females. Each individual made eight pair-wise comparisons in total. For each pair, each respondent had to indicate which of the two cocktails tasted best and which of the two had the most attractive color. The seven cocktails were assigned to the sixty respondents using a balanced incomplete block design. The presentation order effect was eliminated through the use of a linked-pair design, **which ensures that every cocktail is tasted first as often as it is tasted second.** For more details on the design of the cocktail experiment, we refer to Courcoux and Séménou (1997). Just like Courcoux and Séménou (1997), we focus on the taste preference data in our analysis here.

Building on the Bradley-Terry-Luce model for paired comparisons, Courcoux and Séménou (1997) proposed a methodology to determine the number of segments and, for each segment,

Figure 2: Ternary plot showing the experimental region and the seven cocktails used in the study of Courcoux and Séménou (1997).

which of the seven cocktails was preferred. The basic Bradley-Terry-Luce model assumes preference homogeneity among the individuals, but Courcoux and Séménou (1997) embed that model in a latent class framework to cope with the heterogeneity of the individuals' preferences. They estimated their latent class model using an expectation maximization algorithm and determined the number of segments using a test based on Monte Carlo principles. One of the outputs of their analysis is an estimated size for each of the segments identified, similar to the probabilities P_c in the latent class choice model discussed in Section 2.1.3. The final model of Courcoux and Séménou (1997) involves three segments:

- one involving respondents that strongly prefer cocktails with a low proportion of lime (segment 1),
- one in which the respondents are rather indifferent between the seven cocktails, but have a tendency to dislike mango (segment 2), and
- one involving consumers that enjoy the taste of mango (segment 3).

Segment 2 contains the highest proportion of consumers (42%), followed by the first segment and the third, which include 39% and 19% of the respondents, respectively. Table 1 shows the estimated order of preference for the seven cocktails in each of the three segments.

In segment 1, cocktails 4 and 2 have much higher utilities than the other five cocktails, according to Table 1 and Figure 3 in Courcoux and Séménou (1997). For segment 2, the estimated utilities obtained by Courcoux and Séménou (1997) are closer to each other, suggesting that the respondents in this segment do not have strong preferences for one cocktail or another. The respondents in segment 3 have a strong preference for cocktail 1 and dislike all cocktails other than cocktail 4 completely.

Table 1: The estimated order of preference of the seven cocktails in the three segments identified by Courcoux and Séménou (1997).

Segment	Preference
1	$4 > 2 > 1 > 7 > 6 > 5 > 3$
2	$3, 6 > 7 > 2, 5 > 4 > 1$
3	$1 > 4 > 5 > 7 > 6 > 2, 3$

While the approach of Courcoux and Séménou (1997) is certainly elegant and provides much useful information, it can be improved upon by combining state-of-the-art choice models with the traditional models for data from mixture experiments. As a matter of fact, due to the fact that it is based on the Bradley-Terry-Luce model, the analysis method of Courcoux and Séménou (1997) treats the seven cocktails as seven levels of a categorical variable and ignores the composition of the cocktails. The method does not allow the prediction of utilities for mixtures other than the seven used in the experiment, and it does not allow a search for optimal mixtures over the continuous experimental region, i.e., the simplex in Figure 2.

In the next few sections, we extend the analysis of Courcoux and Séménou (1997) in various ways. First, we model the cocktail choices by taking into account the fact that cocktails are mixtures, using the MNLM, the MLM and the LCM. Next, we identify possible consumer segments using the methods described in Section 2.2. So, we use four different two-stage methods based on individual-level preference indicators as well as a latent class mixture-choice model. Third, we create insightful contour plots of the predicted utilities for all possible cocktails and determine the optimal cocktail for each segment we identified. Due to our modeling approach, the optimal cocktails we can identify are different from those actually used in the experiment.

4.2 Estimated mixture-choice models

We started by fitting the three kinds of mixture-choice models we introduced above to the complete set of taste preference data from all sixty respondents, namely the multinomial logit, mixed logit and latent class mixture-choice models. In each of the models, we used the special-cubic Scheffé model for the systematic utility, **because that is the model the simplex-centroid design used by Courcoux and Séménou (1997) is optimal for.** Due to the facts that the cocktail experiment involves three ingredients (so, $q = 3$) and that one of the three linear terms has to be dropped for the choice models to be estimable, the MNLM and MLM models involve six preference parameters. The number of parameters in the LCM models we fitted equals the number of classes times six, plus the number of classes minus one. This

Table 2: Estimates and standard deviations (in parentheses) of the parameters of the MNLM, the MLM and the LCM, along with several goodness-of-fit statistics.

Effect	MNLM	MLM	LCM	
			Class 1	Class 2
Mango	1.1797 (2.47)	0.1043 (2.6)	10.2963 (5.56)	-2.6687 (5.54)
Cassis	-0.6163 (2.22)	-2.3041 (2.35)	5.8104 (5.1054)	-0.9204 (4.8)
Mango*Cassis	5.2216 (8.22)	16.0588 (8.77)	16.8955 (12.75)	-6.7226 (14.86)
Mango*Lime	-4.7861 (10.68)	-2.9624 (11.33)	8.1645 (17.81)	-3.7023 (19.66)
Cassis*Lime	-1.0175 (18.25)	14.1884 (19.13)	19.839 (29.03)	-3.1824 (34.5)
Mango*Lime*Cassis	30.6932 (53.17)	-8.8643 (55.98)	-3.4095 (79.91)	33.7925 (95.45)
Latent class probability			0.649	0.351
-2LL	624.37	581.10	573.24	
AICc	637.96	594.69	603.13	
BIC	648.94	605.67	622.38	

is because the model involves separate preference parameters for each segment, and because the probabilities P_c have to be estimated as well. The linear term we dropped in all our analyses was the one involving the proportion of lime. We estimated the MNLM and the MLM using the JMP Pro 13 software and the LCM using Latent GOLD 5.1, **and we used the corrected Akaike information criterion (AICc) and the Bayesian information criterion (BIC) to quantify the goodness of fit of the various models.**

Table 2 provides the estimates and standard errors of the parameters in the three models. The results reported for the LCM model are those for two segments. This is because, among the fitted LCMs with 1 up to 5 segments, the one involving two segments turned out to have the smallest BIC value. The estimated latent class probabilities show that individuals are more likely to belong to the first segment than to the second segment: the class membership probability for segment 1 is 0.649, while that for segment 2 is only 0.351. Hence, the LCM suggests that the first segment is almost twice as large as the second. The AICc and BIC values for the three models indicate that both the MLM and the LCM fit the data better than the MNLM.

The parameter estimates for the MNLM and for the MLM in Table 2 quantify the preferences of the ‘average’ respondent. The estimates for the LCM quantify the preferences of the ‘average’ respondent in segment 1 and of the ‘average’ respondent in segment 2. As

(a) MNLM

(b) MLM

(c) LCM segment 1

(d) LCM segment 2

Figure 3: The contour plots of the predicted utilities of all cocktails in the experimental region based on the MNLM, the MLM and the two clusters identified by the LCM.

the individual parameters of the Scheffé models are difficult to interpret, we created contour plots that visualize the estimated utility functions. The contour plots are shown in Figure 3, where the predicted utility values produced by the estimated utility functions are represented by different shades of blue. Larger predicted utilities are indicated by darker shades.

Since their upper part has a light color and the upper corner of the experimental region corresponds to cocktails with much lime, the contour plots for the MNLM and the MLM show that the respondents generally prefer cocktails with a small proportion of lime. This is also the case for the first segment identified by the LCM. In contrast, the second segment identified by the LCM has a strong preference for cocktails involving a large proportion of lime. For the MNLM, the maximum predicted utility is achieved by a cocktail that contains 53.64% of mango, 36.36% of cassis and the smallest possible proportion of lime, 10%. For the MLM, the optimal cocktail contains 47.28% of mango, 42.72% of cassis and also the

Table 3: Predicted choice probabilities based on the parameter estimates for the MNLM, the MLM and the LCM in Table 2.

Cocktail	MNLM	MLM	LCM	
			Class 1	Class 2
1	0.1562	0.0916	0.2066	0.0563
2	0.1443	0.1866	0.1418	0.1010
3	0.0593	0.0523	0.0049	0.2687
4	0.2284	0.2818	0.3957	0.0637
5	0.0954	0.0557	0.0468	0.1318
6	0.1401	0.1772	0.0682	0.2342
7	0.1763	0.1548	0.1361	0.1444

smallest possible proportion of lime, 10%. For the first segment identified by the LCM, the optimal cocktail has 57.91% of mango, 32.09% of cassis and 10% of lime (the minimum proportion possible). The second segment’s optimal cocktail includes 30% of mango (the minimum possible), 15% of cassis (the minimum possible) and 55% of lime (the maximum possible). This cocktail corresponds to cocktail 3 in the choice experiment of Courcoux and Séménou (1997). The predicted optima for the MNLM, the MLM and the first segment identified by the LCM are all reasonably close to cocktail 4. Comparing the most and least preferred cocktails in the two segments identified by the LCM with the segments found by Courcoux and Séménou (1997), we can see that there is a match between the LCM’s first and second segment and the first and second segment obtained by Courcoux and Séménou (1997).

Using the parameter estimates in Table 2, we can also calculate predicted choice probabilities for each of the seven cocktails used in the experiment. The seven predicted choice probabilities are shown in Table 3, for the MNLM, the MLM and the two segments identified by the LCM. The choice probabilities produced differ quite substantially, but the first three sets of seven choice probabilities are strongly positively correlated: the correlations lie between 0.77 and 0.90. The choice probabilities for the second segment identified by the LCM are negatively correlated with the other probabilities (correlations lying between -0.39 and -0.76). This confirms that the second segment identified by the LCM involves respondents with preferences that differ substantially from those of the respondents in the other segments. According to the predicted choice probabilities produced by the MNLM, the MLM and the LCM’s first segment, of the original seven cocktails, the fourth (involving 52.5% of mango, 37.5% of cassis and 10% of lime) comes out as the most preferred cocktail. For the LCM’s second segment, cocktail 3 (i.e., the cocktail with the maximum proportion of lime) is the most preferred one.

Comparing our choice probabilities in Table 3 with the choice probabilities reported in Courcoux and Séménou (1997) shows two perfect matches. First, our predicted choice probabilities from the MNLM match those of Courcoux and Séménou (1997) for their analysis with one segment. Second, our predicted choice probabilities from the LCM are identical to those of Courcoux and Séménou (1997) for their analysis with two segments. This perfect match is due to the fact that our models and those of Courcoux and Séménou (1997) contain as many parameters as distinct cocktails, so that the models under study are saturated. For an experimental design involving a larger number of cocktails and/or a mixture-choice model with fewer parameters, there would be no such perfect match.

More importantly, as pointed out above, the AICc and BIC values for the three estimated models indicate that both the MLM and the LCM fit the data better than the MNLM. As a result, there is statistical evidence of preference heterogeneity among the respondents. In the next section, we study this preference heterogeneity in more detail by using individual-level indicator values to identify segments of respondents with the same kinds of preferences. This two-stage segmentation will complement the information produced by the LCM concerning the preference heterogeneity.

4.3 Two-stage segmentation methods

In the previous section, the application of the LCM revealed two segments or classes. In this section, we apply several two-stage segmentation methods. More specifically, we perform a cluster analysis with four sets of indicator values as an input, namely the forces, HB individual-level estimates, Firth individual-level estimates and a combination of these three (i.e., the concatenation of these three). As explained in Section 2.2, we use Ward’s hierarchical clustering method to this end.

4.3.1 The segments identified

The distance graph suggested that, for each set of indicator values we considered, three clusters suffice. In other words, reducing the number of clusters to just two would require uniting sets of respondents with quite different preferences in a single segment. The sizes, gender compositions, most preferred cocktails and least preferred cocktails of the clusters identified based on the four sets of indicator values are summarized in Table 4. The most and least preferred cocktails in the table were determined by evaluating the predicted choice probabilities for the seven cocktails included in the actual choice experiment. For comparison purposes, Table 4 also shows the two clusters identified by the LCM.

Table 4 shows that the sizes of the segments or clusters identified differ from method to

Table 4: Sizes, gender compositions and most/least preferred cocktails of the clusters identified by the two-stage segmentation methods and the LCM.

Method	Forces			HB			Firth			Combination			LCM	
Cluster	1	2	3	1	2	3	1	2	3	1	2	3	1	2
Size	26	24	10	18	28	14	12	40	8	21	20	19	40	20
Male	15	9	6	8	13	9	6	19	5	13	10	7	17	13
Female	11	15	4	10	15	5	6	21	3	8	10	12	23	7
Most preferred	4	4	7	2	4	6	5	4	7	4	7	4	4	3
Least preferred	3	5	5	3	3	1	3	3	5	3	5	3	3	1

method. However, the clusters identified have quite a lot in common:

- Each two-stage segmentation method produces at least one cluster in which cocktail 4 is the most preferred and cocktail 3 is the least preferred. So, each two-stage segmentation method confirms the existence of the first segment identified by the LCM in Section 4.2, as well as that of the first segment identified by Courcoux and Séménou (1997).
- The two-stage segmentation methods all agree that the segment which prefers cocktail 4 and dislikes cocktail 3 represents the largest group of consumers. Also the LCM indicates that this is the largest segment. The LCM, the two-stage segmentation based on Firth individual-level estimates and the two-stage segmentation based on the combined indicator values all suggest that 40 of the 60 respondents have the same most preferred and the same least preferred cocktail.
- Three of the two-stage segmentation methods suggest the existence of a segment in which cocktail 7 is the most preferred, while cocktail 5 is the least preferred. Two of the methods suggest this segment is rather small, while the other suggests it represents a third of the respondents.
- The segments identified by the various methods are all quite gender-balanced.

There are also some differences in the results produced by the four two-stage segmentation methods. For example, the two-stage segmentation based on HB individual-level estimates identifies a segment in which cocktail 6 is the most preferred and cocktail 1 is the least preferred. In this respect, that segment resembles the second segment identified by Courcoux and Séménou (1997). However, none of the other methods produces this kind of segment. It is also noteworthy that none of the segments identified by the two-stage segmentation methods or the LCM resemble the third segment identified by Courcoux and Séménou (1997).

4.3.2 Segment-specific optimal cocktail compositions

A main objective of choice experiments involving mixtures is to determine the optimal mixture for each segment identified, i.e., the mixture that maximizes the predicted utility or the predicted choice probability. Here, the maximization of the utility should be done over the entire simplex region, rather than only over the seven cocktails used in the experiment.

To do so, we first identified the best mixture-choice model for each segment identified. To this end, we estimated 16 different MNL models to the data from all respondents in any given cluster (in doing so, we assume that respondents within a given cluster or segment have homogeneous preferences). The simplest of these 16 models involved the linear terms in the ingredient proportions only, while the most complex model was the special-cubic model including all linear and all interaction terms. For each segment, we used the model with the smallest value for the corrected Akaike information criterion (AICc) to determine the cocktail composition that maximizes the predicted utility and choice probability. As a result, for each segment, we used an all-subset regression approach and used the AICc as a model selection criterion.

The predicted optimal cocktail composition for each of the 14 segments we identified is shown in Figure 4. For example, the predicted optimal cocktail compositions for the three segments identified using the forces are indicated using an inverse triangle, a ‘z’ and an asterisk. For the segment indicated by the inverse triangle, the optimal cocktail involves a mango proportion of 0.56, a cassis proportion of 0.34 and a lime proportion of 0.1. Another segment identified using the forces (indicated by the ‘z’) prefers a cocktail involving a mango proportion of 0.48, a cassis proportion of 0.42 and a lime proportion of 0.1. The final segment identified using the forces and indicated by the asterisk prefers a cocktail with a mango proportion of 0.44, a cassis proportion of 0.32 and a lime proportion of 0.24.

Figure 4 shows that, for each of the five segmentation methods we studied, at least one predicted optimum is situated quite centrally on the horizontal axis. This implies that the predicted optimal cocktail composition involves a minimum of lime (i.e., a lime proportion of 0.1), a proportion of mango between 0.45 and 0.6, and a proportion of cassis between 0.3 and 0.45. Most other predicted optima involve a proportion of lime smaller than 0.3, a proportion of mango slightly large than 0.4 and a proportion of cassis slightly larger than 0.3. Only one predicted optimum involves a large proportion of lime: for one segment identified by the LCM, the predicted optimal cocktail composition uses the maximum proportion of lime of 0.45 and the minimum proportions of 0.3 and 0.15 for mango and cassis, respectively.

Figure 4 also visualizes the agreement and disagreement between the various segmentation methods concerning the optimal cocktail composition:

Figure 4: Cocktail compositions that are optimal in terms of predicted taste preference for the 14 segments identified by the various segmentation methods.

- Seven of the 14 predicted optima involve the minimum proportion of lime allowed. These predicted optima appear on the bottom axis of the ternary plot in Figure 4.
- Four of the remaining predicted optima are located quite closely to the centroid of the experimental region, corresponding to proportions of 0.25, 0.45 and 0.35 for lime, mango and cassis, respectively.
- Three of the predicted optima deviate substantially from all others. One of these corresponds to a segment identified using the LCM. Another corresponds to a segment identified using HB individual-level estimates, and the final one corresponds to a segment obtained from Firth individual-level estimates.

The optimal cocktail compositions shown in Figure 4 are point predictions, and ignore the uncertainty associated with the estimates of the model parameters and the resulting uncertainties in the predicted utility and choice probability values. To investigate the uncertainty about the predicted optimal cocktail compositions, we randomly drew 1,000 parameter vectors from the approximate multivariate normal distribution of the parameter estimates. For each randomly drawn parameter vector, we determined the corresponding predicted optimum. Visualizing the resulting 1,000 predicted optima in a ternary plot then provides a good picture of the prediction uncertainty.

We visualize the standard deviations of the ingredient proportions of the 1,000 predicted optima for each segment identified by each of the five segmentation procedures in Figure 5. Obviously, the smaller the standard deviations for any given segment, the more confidence

Figure 5: Standard deviation of ingredients for 1,000 simulated optimal taste cocktails.

we can have in the predicted optimum. For all segments, except those labeled HB 1, Firth 1, Firth 3 and Combination 1, the standard deviations are rather low, indicating that the predicted optima in Figure 4 are generally trustworthy. For the first segment identified using HB individual-level estimates (HB 1), the first and third segment identified using Firth individual-level estimates (Firth 1 and 3) and the first segment based on the combination of the forces, the HB individual-level estimates and the Firth individual-level estimates (Combination 1), the standard deviations are thus quite large. This suggests that small changes in the parameter estimates would result in very different predicted optima. Therefore, the predicted optima for these segments should be interpreted with more care. Figure 6 shows two extreme cases: there is much uncertainty about the predicted optimal cocktail composition for the first segment identified using Firth individual-level estimates, while there is little such uncertainty for the third segment based on HB individual-level estimates.

The smallest standard deviations are for the segments labeled Forces 2, HB 2, HB 3, Firth 2, Combination 2, LCM 1 and LCM 2. For these segments, there is little uncertainty concerning the composition of the optimal cocktail. This implies that we can be confident that there is a group of respondents that prefer cocktails with the smallest possible proportion of lime and between 45% and 60% of mango, and that there is another group that prefers cocktails with about 25% of lime, 45% of mango and 30% of cassis. In other words, one group of respondents has a preference for a cocktail that resembles cocktail 4 used in the actual taste experiment and shown in Figure 2. The other group prefers a cocktail that resembles cocktail 7 from the actual taste experiment.

Based on the models with the lowest AICc values for each of the segments identified, we

(a) Firth 1

(b) HB 3

Figure 6: Ternary plots visualizing the uncertainty about the predicted optimal cocktail compositions for two of the segments identified.

Table 5: Predicted choice probabilities for each cocktail actually used in the experiment for each cluster or segment, **including those obtained by Courcoux and Séménou (1997). The latter are indicated by the abbreviation C&S.**

Cluster	Cocktail						
	1	2	3	4	5	6	7
Forces 1	0.1419	0.0646	0.0298	0.3098	0.1822	0.1734	0.0982
Forces 2	0.1341	0.2040	0.0540	0.2845	0.0532	0.1028	0.1674
Forces 3	0.1197	0.1957	0.1183	0.0695	0.0280	0.0618	0.4071
HB 1	0.1699	0.3206	0.0782	0.1867	0.0825	0.0811	0.0810
HB 2	0.1913	0.1256	0.0339	0.3406	0.0470	0.0746	0.1871
HB 3	0.0306	0.0399	0.0513	0.0696	0.1115	0.3579	0.3392
Firth 1	0.1827	0.0602	0.0287	0.1617	0.3180	0.1827	0.0660
Firth 2	0.1242	0.1427	0.0457	0.2991	0.0753	0.1415	0.1714
Firth 3	0.1990	0.1936	0.1929	0.0825	0.0202	0.0351	0.2768
Combination 1	0.1354	0.0494	0.0283	0.2851	0.2016	0.2007	0.0994
Combination 2	0.1247	0.1377	0.1650	0.1081	0.0451	0.1386	0.2808
Combination 3	0.1289	0.3378	0.0066	0.3949	0.0293	0.0304	0.0721
LCM 1	0.2430	0.1744	0.0040	0.2907	0.0522	0.0741	0.1616
LCM 2	0.0505	0.0860	0.3198	0.0659	0.1271	0.2233	0.1274
C&S 1	0.0920	0.3380	0.0000	0.4770	0.0080	0.0250	0.0610
C&S 2	0.0750	0.1190	0.2150	0.0930	0.1110	0.2160	0.1720
C&S 3	0.4620	0.0000	0.0000	0.2630	0.1250	0.0520	0.0970

also ranked the seven cocktails used in the actual taste experiment in terms of the predicted choice probabilities. These predicted choice probabilities are shown in Table 5, **along with the choice probabilities reported by Courcoux and Séménou (1997) for the three segments their analysis suggested**. We used the rankings to investigate which of our segments match those of Courcoux and Séménou (1997) best. To this end, for each of the 42 combinations of one of our 14 segments and one of the three segments of Courcoux and Séménou (1997), we computed the rank correlation coefficient. The rank correlations show that the first segment of Courcoux and Séménou (1997) is very similar to those labeled LCM 1 (0.96), Forces 2 (0.93), HB 1 (0.93) and Firth 2 (0.82) in our analysis. The second cluster of Courcoux and Séménou (1997) shows a strong similarity with the segment labeled LCM 2 in our analysis. Lastly, the third cluster of Courcoux and Séménou (1997) shows some similarity with the segments labeled Forces 1 (0.70) and Firth 1 (0.74). The last result to some extent came as a surprise because the third segment of Courcoux and Séménou (1997) likes cocktail 1 most and cocktails 2 and 3 least and, as indicated by Table 4, none of our segments had cocktail 1 as most preferred cocktail and cocktail 2 or 3 as least preferred cocktail. This shows that considering a full ranking of multiple cocktails sheds some additional light on the similarity or dissimilarity of segments.

5 Color preference

While the taste preference data studied in detail in Section 4 are the most important, it is also interesting to study the preferences of the 60 respondents when it comes to the color of the cocktail. Therefore, for each of the 14 segments we identified in Section 4, we determined the best MNLM for the color preference data. To this end, we again fitted 16 different models for each segment identified, and selected the one with the smallest AICc value to predict the optimal cocktail in terms of color. The 14 resulting predicted optimal cocktail compositions are shown in Figure 7. Comparing Figure 4, showing the predicted optima for taste, and Figure 7, containing the predicted optima for color, we observe the following patterns:

- For five of the seven segments that prefer the taste of a cocktail with the minimum proportion of lime, the cocktail preferred in terms of color also has a minimum a lime. The proportions of mango and cassis in the optima for taste and color are quite similar as well. So, for these segments, there is one cocktail composition that approximately optimizes both the taste and the color.
- For two of the five segments that prefer the taste of a cocktail with a lime proportion of 0.25–0.30, the preferred cocktail in terms of color has a minimum of lime and roughly

Figure 7: Cocktail compositions that are optimal in terms of predicted color preference for the 14 segments identified by the various segmentation methods based on the taste preference.

equal proportions of mango and cassis. For these two segments, the cocktail that optimizes the taste is different from the one that is most preferred in terms of color, but not extremely so.

- For all other segments (especially for LCM 2 and Firth 1), the preferred cocktail in terms of color involves as little mango as possible, quite a large proportion of cassis and a quite small proportion of lime. For some of these segments, the cocktail that optimizes the taste is similar to the one that is least preferred in terms of color.

One possibility to reconcile the preferences for taste and color in each segment identified is to perform an optimization of the two responses simultaneously. To this end, we used the desirability function approach described by Derringer and Suich (1980) and implemented in JMP Pro 13. The resulting optimal cocktails are shown in Figure 8. We observe five distinct groups:

- For six of the segments (labeled Forces 1, Forces 2, HB 2, Firth 2, Combination 3 and LCM 1), the cocktail that jointly optimizes the taste and the color involves the minimum proportion of lime, about 50% of mango and about 40% of lime.
- For the segments labeled Forces 3 and HB 1, the optimal cocktail contains mango, cassis and lime proportions of 0.3, 0.6 and 0.1, respectively. That cocktail corresponds to the bottom right-hand corner of the experimental region.

Figure 8: Cocktail compositions that optimize the predicted taste and color preferences simultaneously for the 14 segments identified by the various segmentation methods based on the taste preference.

- For the segments labeled LCM 2, Firth 1 and Combination 1, the optimal cocktail contains mango, cassis and lime proportions of 0.3, 0.4–0.43 and 0.27–0.3, respectively.
- For the segments labeled HB 3, Firth 3 and Combination 2, the optimal cocktail contains 40–43% of mango, 38–40% of cassis and 17–21% of lime.

Comparing the results for the joint optimization of taste and color preference and those for optimizing the taste only, we can see that, for a substantial number of the segments we identified with the various segmentation methods, there is substantial agreement concerning the optimal cocktail composition. Producing one cocktail with the minimum proportion of lime, about 50% of mango and about 40% of cassis and another one with roughly 20% of lime, 40% of mango and 40% of cassis will please nine of the 14 segments we identified quite a lot, both in terms of taste and in terms of color. This can be seen by comparing Figures 4 and 8.

6 Discussion

This paper has three main contributions. First, we combined commonly used choice models with traditional Scheffé models for data from mixture experiments. Second, we compared five different ways to identify consumer segments. Third, we applied the newly obtained model to data from an actual choice experiment involving mixtures.

For our application, the aggregate MNLM (i.e., the MNLM applied to the full data set) is not appropriate to model the choices made, due to the fact that the respondents have heterogeneous preferences. We modeled the heterogeneity using the MLM and the LCM, and we also applied the MNLM to the data from each individual separately. Even though the five segmentation methods we used do not completely agree on the number of segments and on the preference of the respondents in the various segments, the results produced strongly suggest the existence of two main groups of consumers that share the same preferences in terms of taste.

In general, we recommend to try different segmentation methods in parallel. This is because, for each individual respondent in a taste experiment, very few data points are available, which leads to imprecise individual-level information. We experienced this, for instance, when acquiring HB individual-level estimates both in the JMP Pro 13 software and in SAS 9.4. We would therefore be hesitant to rely on one segmentation method only. We would generally also try to incorporate covariate information concerning the respondents in the choice models fitted. This is because we like to detect any possible sources of the preference heterogeneity. In the application we studied, the only covariate available in the data set was the gender, but this covariate was insignificant in each of the models we fitted.

Finally, we would like to point out that much research can still be done on the optimal design of choice experiments involving mixtures. Courcoux and Séménou (1997) cleverly combined existing experimental designs to set up their choice experiment. More specifically, they combined three well-known standard experimental designs, one for mixtures (a simplex-centroid design), one for blocked experiments (a balanced incomplete block design) and one for paired comparisons (the linked-pair design). Assuming preference homogeneity, Ruseckaite et al. (2017) showed how the optimal experimental design framework can be used to create D-optimal designs for choice experiments with mixtures. However, as the goal of such choice experiments generally is to optimize the mixture, computing I-optimal designs (which minimize the average variance of prediction over the entire experimental region) might be a better alternative. Moreover, rather than assuming preference homogeneity when creating a design, it would make sense to seek designs that perform well in the presence of respondent heterogeneity. Finally, Ruseckaite et al. (2017) ignored a possible order effect. By including a within-pair order effect in the choice models introduced here, as described by Goos and Grossmann (2011), the analysis and optimal experimental design approach can be modified to account for possible order effects as well. Another useful extension of our work would be to combine mixture-process variable models with choice models. This is useful to study preferences for mixtures that require extra processing. For instance, to model preferences for bread, not only the composition of the dough mixture is relevant, but also the baking

temperature and the baking time.

Acknowledgement

We would like to thank Prof. Dr. Courcoux for making the data from Courcoux and Séménou (1997) available to us. We also acknowledge the financial support of the Fund for Scientific Research Flanders (FWO-Vlaanderen).

References

- Atwood, C. L. (1969). Optimal and efficient designs of experiments. *The Annals of Mathematical Statistics*, 40(5):1570–1602.
- Ben-Akiva, M. and Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 12(1):9–24.
- Bennett, J. and Blamey, R. (2001). *The Choice Modelling Approach to Environmental Valuation*. Edward Elgar Publishing.
- Cornell, J. A. (2011). *Experiments with mixtures: designs, models, and the analysis of mixture data*. John Wiley & Sons.
- Courcoux, P. and Séménou, M. (1997). Une méthode de segmentation pour l’analyse de données issues de comparaisons par paires. *Revue de Statistique Appliquée*, 45(2):59–69.
- Crabbe, M., Jones, B., and Vandebroek, M. (2013). Comparing two-stage segmentation methods for choice data with a one-stage latent class choice analysis. *Communications in Statistics: Simulation and Computation*, 42(5):1188–1212.
- Derringer, G. and Suich, R. (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology*, 12(4):214–219.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.
- Goos, P. and Grossmann, H. (2011). Optimal design of factorial paired comparison experiments in the presence of within-pair order effects. *Food Quality and Preference*, 22(2):198 – 204.
- Goos, P. and Jones, B. (2011). *Design of Experiments: A Case-Study Approach*. New York: Wiley.

- Goos, P., Jones, B., and Syafitri, U. (2016). I-optimal design of mixture experiments. *Journal of the American Statistical Association*, 111(514):899–911.
- Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8):681–698.
- Hensher, D. A. and Johnson, L. W. (2018). *Applied Discrete-Choice Modelling*. Routledge.
- Hensher, D. A., Rose, J. M., and Greene, W. H. (2005). *Applied Choice Analysis: A Primer*. Cambridge University Press.
- Kamakura, W. A. and Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 26(4):379–390.
- Kessels, R., Jones, B., and Goos, P. (2019). Using Firth’s method for model estimation and market segmentation based on choice data. *Journal of Choice Modelling*, 31:1–21.
- Kiefer, J. (1961a). Optimum designs in regression problems, II. *The Annals of Mathematical Statistics*, pages 298–325.
- Kiefer, J. (1961b). Optimum experimental designs V, with applications to systematic and rotatable designs. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 381–405. University of California Press.
- Kupfer, F., Kessels, R., Goos, P., Van de Voorde, E., and Verhetsel, A. (2016). The origin-destination airport choice for all-cargo aircraft operations in Europe. *Transportation Research Part E: Logistics and Transportation Review*, 87:53–74.
- Lancsar, E. and Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics*, 26(8):661–677.
- Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge University Press.
- Louviere, J. J. and Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of Marketing Research*, 20(4):350–367.
- Luyten, J., Kessels, R., Goos, P., and Beutels, P. (2015). Public preferences for prioritizing preventive and curative health care interventions: A discrete choice experiment. *Value in Health*, 18(2):224–233.

- McFadden, D. (1986). The choice theory approach to market research. *Marketing Science*, 5(4):275–297.
- McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.
- Nylund, K. L., Asparouhov, T., and 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*, 14(4):535–569.
- Punj, G. and Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20(2):134–148.
- Ruseckaite, A., Goos, P., and Fok, D. (2017). Bayesian D-optimal choice designs for mixtures. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(2):363–386.
- Scheffé, H. (1958). Experiments with mixtures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2):344–360.
- Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25(2):235–263.
- Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics*, 41(22):2915–2924.
- Train, K. (2001). A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit. *University of California, Berkeley*, pages 1–13.
- Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Vardakis, M., Goos, P., Adriaensen, F., and Matthysen, E. (2015). Discrete choice modelling of natal dispersal: Choosing where to breed from a finite set of available areas. *Methods in Ecology and Evolution*, 6(9):997–1006.