Cover page

(This page is intentionally left blank)

New Flexible Models and Design Construction Algorithms for Mixtures and Binary Dependent Variables

New Flexible Models and Design Construction Algorithms for Mixtures and Binary Dependent Variables

Nieuwe flexibele modellen en algoritmen voor het ontwerp van experimenten voor mengsels en binaire afhankelijke variabelen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on

Thursday, January 12, 2017 at 13:30 hours

by

AISTE RUSECKAITE

born in Vilnius, Lithuania

Erasmus University Rotterdam

Doctorate Committee

Promotors:	Prof.dr. D. Fok
	Prof.dr. P. Goos
Other members:	Prof.dr. B. Donkers
	Prof.dr. S. Hess
	Prof.dr. R. Paap

ISBN: 978 90 3610 471 5

This book is no. 670 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Acknowledgments

I would like to express my sincere gratitude to my doctoral promotors Professor Dennis Fok and Professor Peter Goos. For sharing all the knowledge, for keeping me positive and inspired, and, most importantly, for being patient. If not them, this thesis would not have been possible. ”Great men are not born great, they grow great” (M. Puzo, *The Godfather*), and it is much easier to grow great with great teachers. Thank you.

Contents

Chapter 1

Introduction and Outline

1.1 Introduction

In many regression settings, the explanatory variables involve a *mixture* of *ingredients* whose proportions sum to one. As an example, consider a yoghurt, which can be seen as a mixture of milk, fruit puree and maple syrup. In a transportation setting, one can recognize a mixture of different components of travel time (e.g. in-vehicle and out-of-vehicle travel time). In marketing, advertising campaigns can be described as mixtures of different advertising media (e.g. TV, magazine and newspaper advertising). From the examples above, it is clear that many products and services are in fact mixtures. Therefore, it is surprising that there exist a number of voids in the literature for mixtures. I begin this thesis with the aim of filling them.

One way to obtain mixture data is by running an experiment and recording the outcome values for each mixture in the experiment. Running experiments is costly, therefore, efficient experiments are desired. Typically, in experiments, respondents are asked to rate or rank products or services presented to them. Such experiments become less appealing if the number of alternatives that has to be considered is large and when the differences between alternatives are small and subjective. In these scenarios, choice experiments, where respondents are asked to select the product or service that they like best from a group of alternatives, are preferred. However, to date, there are no studies on how to optimally design choice experiments for products or services that are mixtures. For this reason, this thesis begins by developing methods for constructing optimal choice experimental designs for mixtures.

In many scenarios, only mixture proportions have an effect on the response. In other cases, the total amount of the mixture matters as well. Consider advertising campaigns that can be described as mixtures of advertisements in multiple advertising media. Not only the proportions of the total advertising budget invested in each medium are important to determine the impact of the campaign (e.g. 30% of the total advertising budget invested in TV advertising and 70% on the Internet) but also the total advertising budget itself. I refer to this total advertising budget variable as the total amount. This kind of data is called mixture-amount data. The models that exist for mixture-amount data date back to the 1980s and have several drawbacks which limit their usefulness for these data. Therefore, the third chapter of this thesis develops new flexible models for mixture-amount data.

The last chapter is somewhat different from the other two. It is based on a revealed preference case study of consumer attitudes with respect to electric and hybrid vehicles. Electric and hybrid vehicles have been on the market for many years. However, despite strong advertising and financial incentives, their market penetration is still limited. Therefore, many researchers in transportation continue developing advanced mathematical or statistical methods and data collection protocols to explain what prevents consumers from choosing environmentally friendly vehicles.

The vehicle attributes when modeling choices of electric or hybrid vehicles are often the same or rather similar: range, accessibility to charging stations, operating costs and the price. With respect to individual characteristics, more and more researchers recognize the importance of consumers' environmental attitudes for explaining the choice. Additionally, a number of researchers emphasize the fact that attitudes and opinions greatly differ among individuals. However, to date, there are no studies that account for that beyond individual-specific dummy variables. Therefore, with the last chapter of my thesis, I aim to contribute to this stream of research by introducing a new model that allows for a heterogenous impact of the parameter of interest across individuals.

1.2 Outline

This thesis consists of three self-contained chapters that can be read independently. In this section, I introduce each of them in more detail.

Chapter 2 is based on Ruseckaite et al. (2016b). Here, I introduce mixture models in the choice context and develop an optimal design construction algorithm for choice experiments involving mixtures. Choice experiments may help to determine how a respondent's choice of a product or service is affected by the combination of ingredients. In such experiments, individuals are confronted with sets of hypothetical products or services and they are asked to choose the most preferred product or service from each set. Choice experiments are costly, and therefore it is important to design them efficiently. However, there exist no studies on the optimal design of choice experiments involving mixtures. For this reason, Chapter 2 develops a method for generating optimal designs for such choice experiments. I present two algorithms for obtaining optimal designs involving mixtures, namely, a particle swarm optimization algorithm and a mixture coordinate-exchange algorithm. I demonstrate the large increase in statistical efficiency if such an optimal design is used.

As a motivating application, I consider the mixture experiment described by Courcoux and Séménou (1997). In this experiment, preference information on fruit cocktails was elicited. There were seven fruit cocktails in the experiment which were made of mango juice, blackcurrant syrup and lemon juice, in different proportions. In the experiment, 60 respondents had to taste eight pairs of cocktails and to indicate each time the cocktail they preferred. In the study by Courcoux and Séménou, an ad-hoc experimental design was used. I compare this ad-hoc experimental design used by Courcoux and Séménou (1997) to an optimal experimental design built using the approach introduced in Chapter 2.

In many scenarios, only the mixture proportions matter for the outcome variable. In such cases, mixture models suffice. In other scenarios, the total amount of the mixture matters as well. In these cases, one needs mixture-amount models. As an example, consider advertisers who have to decide on the advertising media mix (e.g. 30% of the expenditures on TV advertising, 10% on radio and 60% on online advertising) as well as on the total budget of the entire campaign. To model mixture-amount data, the current strategy is to express the response in terms of the mixture proportions and specify mixture parameters as parametric functions of the amount. The reason behind this approach is that if the total amount of the mixture affects the impact of mixture proportions, the parameters corresponding to the mixture ingredients in a model need to vary with the amount. However, such models require the specification of a functional form relating the mixture parameters to the amount variable a priori. Correctly specifying such a function is not straightforward. Furthermore, when a flex-

ible parameterization is used, there are many parameters to estimate. Therefore, in Chapter 3, which is based on Ruseckaite et al. (2016a), I introduce a new modeling approach which is flexible but parsimonious in the number of parameters. The model is based on so-called Gaussian processes and avoids the necessity to a priori specify the shape of the dependence of the mixture parameters on the amount. I also demonstrate the model's added value when compared to standard models for mixture-amount data. I provide two illustrations. The first one deals with the reaction of mice to mixtures of hormones when these are administered in different amounts. The second one concerns the recognition of advertising campaigns. The mixture here is the particular media mix (TV and magazine advertising) used for a campaign. As the total amount variable, I consider the total advertising campaign exposure. The dependent variable here is binary and indicates whether an advertising campaign is recognized (1) or not (0).

Chapter 4 is based on the project on which I worked together with M. Bierlaire during my research visit to the TRANSP-OR lab at École Polytechnique Fédérale de Lausanne (EPFL). In this chapter, using revealed preference data on green vehicle purchases in France in years $2010-2014$, I develop a new choice model that accounts for latent environmental consciousness, where environmental consciousness is allowed to have a heterogeneous impact on the vehicle choice across the population. I associate this impact to an individual's age. However, any other individual characteristic could be used instead. Technically, I assume that the impact of environmental consciousness is a function of one's age. A priori, I do not know what such dependence might look like. Instead of assuming some functional form, I apply a new approach based on the Gaussian process prior, to infer this dependence from data. As environmental consciousness is not directly observed, I infer it for each individual in the sample from the observed attitudinal indicators. I also account for technical and price attributes of vehicles. However, I pay special attention to understanding whether and how environmental consciousness drives choices regarding green vehicles. I demonstrate the added value of the new approach if compared to the model where the parameter of environmental consciousness is assumed to be constant for all individuals.

Chapter 2

Bayesian D-Optimal Choice Designs for Mixtures

2.1 Introduction

Choice experiments are commonly used to obtain information on consumer preferences. In such experiments, products or services are typically characterized by combinations of attribute levels called profiles or alternatives. Respondents select the most preferred alternative from a group of alternatives called a choice set. They repeat this task for several other choice sets presented to them. All choice sets together make up the experimental design (Kessels et al., 2009; Rose and Bliemer, 2009). The preferences recorded in the course of a choice experiment allow us to estimate the importance of each attribute and its levels. Such estimates are crucial to successfully design new products and services, to predict market shares, and to determine willingnesses to pay.

Alternative ways to quantify consumer preferences utilize rating and/or ranking approaches. These approaches are, however, not appealing if the number of alternatives that has to be considered is large and when the differences between alternatives are small and subjective. In such scenarios, many researchers in sensometrics, marketing, transportation, environmental and health economics, and psychology have resorted to pairwise comparisons (David, 1963; Agresti, 2002). In pairwise comparison studies, the respondents repeatedly evaluate pairs of alternatives, which is cognitively less demanding and considered to be more reliable than rat-

ing or ranking large sets of alternatives. Paired comparison studies can be viewed as choice experiments with choice sets of two alternatives.

While the design of choice experiments involving categorical and (general) quantitative attributes has received substantial attention in the literature, the design of choice experiments where the attributes are proportions of ingredients in a mixture has not been studied at all. This is surprising given the fact that many consumer products are in fact mixtures and given the fact that there are many examples of experiments involving mixtures in the literature. For instance, Cornell (2002) describes several experiments where respondents have to rate or rank mixtures. The examples in Cornell's textbook deal with sports drinks (with various sweeteners as ingredients), fish patties (with mullet, sheepshead and croaker as ingredients) and tobacco blends (with flue-cured tobacco, burley, Turkish blend and processed tobacco as ingredients), among other things. Sahrmann et al. (1987) describe a mixture experiment involving an alcoholic drink based on orange juice, vodka and an Italian liquor - Galliano. Finally, Rehman et al. (2007) determine the optimal blend of two British wheat varieties, Mercia and Galahad, for making chapatti (a flat bread that is a staple food in Pakistan, India and certain parts of Africa).

In this chapter, we take as a leading example the mixture experiment as described in Courcoux and Séménou (1997). In this experiment, preference information on fruit cocktails is elicited. The authors considered seven fruit cocktails that are made of mango juice, blackcurrant syrup and lemon juice, in different proportions. In the experiment, 60 respondents had to taste eight pairs of cocktails and, for each pair, they had to indicate which one they preferred. For each respondent, the pairs were formed by selecting two out of seven different cocktails, using an ad-hoc experimental design.

It is clear that there are many products that can be considered as mixtures of ingredients and that choice-based experiments are often used. However, to date, there are no studies describing how to efficiently design choice-based mixture experiments. In this chapter, we show how to optimally design choice experiments involving mixtures. We bring together the large body of work on the design and analysis of mixture experiments and the work on the optimal design of discrete choice experiments, also known as stated preference studies. The focus in this chapter is on applications in sensometrics. However, our work is also very relevant for other application areas. For instance, in transportation, travel time can be viewed as a mixture of congested travel time and free-flow travel time and the total cost may be a

mixture of fuel cost and tolls. In the RGB color model, colors are represented as mixtures of the ingredients red, green and blue. The preference for the design of company logos could therefore also be framed as a choice-based mixture experiment.

In the next section, we review the relevant literature on the design of choice experiments and mixture experiments. In Section 2.3, we discuss how we incorporate mixtures in the multinomial choice model and present the design optimality criterion. We present two algorithms for obtaining optimal designs involving mixtures in Section 2.4. We show a selection of our computational results in Section 2.5, starting with locally optimal designs and ending with Bayesian optimal designs. We pay special attention to the performance of the newly developed designs relative to utility-neutral designs, a class of locally optimal choice designs that has received much attention in the literature. In Section 2.6, we compare the ad-hoc experimental design used by Courcoux and Séménou (1997) for the fruit cocktail study to an optimal experimental design we built using the approach introduced in this chapter. Section 2.7 contains a conclusion and discussion.

2.2 Literature

This chapter combines two streams of literature, that on the optimal design of choice experiments and that on the design of mixture experiments. In this section, we describe both streams in turn.

The first paper to discuss the D-optimal design of choice experiments was Huber and Zwerina (1996). They show that the optimal design depends on the true preference structure, or, in other words, the parameters of the choice model. They solve this circular problem by constructing the choice design for one single a priori chosen value for the model parameters. This yields the so-called locally optimal design. Sándor and Wedel (2001) generalize this to allow for uncertainty about the parameter values. The resulting Bayesian optimal design approach relies on a proper prior distribution for the parameters. Since 2001, the Bayesian optimal design approach has received much attention in the choice design literature. More recent work in this area has focused on applications in transportation (Hensher and Rose, 2009) and environmental economics (Vermeulen et al., 2011), fast approaches to evaluate the Bayesian optimality criteria (Bliemer et al., 2008; Yu et al., 2010), models other than the multinomial logit model (Bliemer et al., 2009; Yu et al., 2009; Goos et al., 2010; Bliemer

and Rose, 2010; Yu et al., 2011) or design approaches that take into account the complexity of choice tasks (Kessels et al., 2011a; Danthurebandara et al., 2011).

Some authors use a very specific locally optimal design (the so-called utility-neutral design) for their choice experiments, i.e., they assume that all model parameters equal zero. This assumption implies that all alternatives have the same utility and respondents are therefore indifferent between all alternatives. We refer to Grasshoff et al. (2004, 2003); Grossmann et al. (2006, 2009); Burgess and Street (2005); and Street and Burgess (2007) for D-optimal designs derived under this assumption.

The research on mixture experiments has focused on industrial and bioscience applications, where linear regression models are used and experiments are usually completely randomized or blocked (Cornell, 2002; Smith, 2005). D-optimal completely randomized designs for the most commonly used linear mixture models and experimental regions have been known for many years (Kiefer, 1961; Uranisi, 1964). To construct mixture experiments for more general model types and experimental regions, Piepel et al. (2005) introduced a mixture coordinate-exchange algorithm. Wong et al. (2015) adapted the particle swarm optimization (PSO) algorithm for the same purpose. The PSO algorithm itself was originally introduced in the optimal design literature by Chen et al. (2011). The original coordinate-exchange algorithm is due to Meyer and Nachtsheim (1995).

In spite of all published work on the design of choice experiments and the design of mixture experiments, the construction of optimal choice designs for mixtures is still an unexplored research area. In this chapter, we fill this void in the literature and investigate the performance of the coordinate-exchange and the PSO algorithms for this purpose. We focus on these two kinds of algorithms because both of them have been proven successful when generating designs for mixture experiments outside the choice context (see Piepel et al. (2005); Wong et al. (2015)).

2.3 Models and design criterion

In this section, we first describe general models for data from mixture experiments. Next, we extend these ideas to the choice context and more specifically to the multinomial logit model. We also discuss the design selection criterion that results from this approach.

In a mixture choice experiment, an *alternative* (product or service) is represented by a combination of q *ingredient proportions*, $x_i,\ i = 1, \ldots, q$, where $x_i \geq 0$ and $\sum_{i=1}^{q} x_i = 1$. The group of alternatives presented to a respondent in a single task is called a *choice set*. The collection of all choice sets is called the *design*. We denote the total number of alternatives in an experimental design by n, the number of choice sets by S and the number of alternatives within a choice set by J, so that $n = SJ$.

2.3.1 Models for data from mixture experiments

In order to model preference as a function of mixtures of ingredients, a suitable model is required, i.e., a model that can handle proportions as explanatory variables. Such models have been developed for situations involving a continuous response y. As a result, they are framed in the context of linear regression. The *mixture constraint* defined by $\sum_{i=1}^{q} x_i = 1$ has a substantial impact on the models that can be considered. The first major consequence of the mixture constraint is that a linear regression model for mixture data cannot contain an intercept. Moreover, cross-products $x_i x_j$ (generally required for quantifying interaction effects) and squares x_i^2 (generally required for quantifying quadratic effects) cannot be simultaneously included in the model, as this would lead to perfect collinearity. As a result, the mixture constraint naturally leads to the family of models proposed by Scheffé (1958, 1963). The first-order Scheffé model for a continuous dependent variable y is given by

$$y = \sum_{i=1}^{q} \beta_i x_i + \varepsilon, \tag{2.1}$$

whereas the second-order Scheffé model is given by

$$y = \sum_{i=1}^{q} \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} x_i x_j + \varepsilon. \tag{2.2}$$

The so-called special-cubic model can be written as

$$y = \sum_{i=1}^{q} \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} x_i x_j + \sum_{i=1}^{q-2} \sum_{j=i+1}^{q-1} \sum_{k=j+1}^{q} \beta_{ijk} x_i x_j x_k + \varepsilon. \tag{2.3}$$

The coefficient β_i in Equations (2.1)-(2.3) can be interpreted as the expected response if a product consists of ingredient i only, that is, if $x_i = 1$. We cannot interpret this coefficient as the partial effect of ingredient i, since changing x_i requires at least one other proportion to be changed as well. Therefore, it is relatively difficult to interpret parameters in the Scheffé models.

If we expect interaction effects like synergism (interaction of ingredients such that the total effect is greater than the sum of the individual effects) or antagonism (interaction of ingredients such that the total effect is smaller than the sum of the individual effects), we should use the second-order or special-cubic model in Equation (2.2) or (2.3). As the number of terms in these Scheffé models increases rapidly with the number of ingredients q, utilizing these models requires a large number of observations.

In an unconstrained mixture experiment, the experimental region is the set of all possible combinations of the proportions $x_1, x_2, \ldots, x_q$ that satisfy the mixture constraint. That region is a $(q-1)$-dimensional regular simplex, which is an equilateral triangle when $q = 3$. If we impose constraints on the ingredient proportions, such as lower and/or upper bounds, we obtain a constrained experimental region. When only lower bounds are used, the experimental region remains simplex-shaped and it is in this case common to redefine the proportions in terms of so-called L-pseudo components, each of which again takes values on the interval $[0, 1]$ and the sum of which is one. If we denote the lower bound for ingredient i as L_i, where $0 \leq L_i \leq x_i$, then the ith pseudo component x_i' is defined as $x_i' = \frac{x_i - L_i}{1-L}$, where $L = \sum_{i=1}^{q} L_i < 1$. After this transformation, one is left with a standard mixture model and experimental region in terms of the pseudo components.

2.3.2 Multinomial logit model

To estimate a preference structure and to model the choices made in mixture choice experiments, we build on the multinomial logit (MNL) model. The model is based on random utility theory and expresses the utility of alternative j in choice set s, denoted by u_{js}, as a function of the observed alternative specific attributes plus an error term, that is, $u_{js} = \mathbf{f}(\mathbf{x}_{js})'\boldsymbol{\beta} + \varepsilon_{js}$, where $\mathbf{x}_{js}$ is a vector that contains the q proportions corresponding to this alternative j in choice set s, $\mathbf{f}(\mathbf{x}_{js})$ represents the model expansion of these attributes and $\boldsymbol{\beta}$ is the corresponding parameter vector. In each choice set, a respondent chooses the

alternative that has the largest utility. The probability that a respondent chooses alternative j in choice set s is therefore

$$p_{js} = P[u_{js} > \max\{u_{1s}, \ldots, u_{j-1,s}, u_{j+1,s}, \ldots, u_{Js}\}].$$

In the MNL model, the error terms ε_{js} are assumed to be mutually independent and to follow the so-called log Weibull distribution (also known as type-I extreme value distribution). Then, the probability that a respondent chooses alternative j in choice set s can be written as

$$p_{js} = \frac{\exp(\mathbf{f}(\mathbf{x}_{js})'\boldsymbol{\beta})}{\sum_{t=1}^{J} \exp(\mathbf{f}(\mathbf{x}_{ts})'\boldsymbol{\beta})},$$

see Ben-Akiva and Lerman (1985) for a detailed derivation.

In the mixture context, the attributes are ingredient proportions. Therefore, the model expansions we use in this chapter are based on, though not identical to, the basic Scheffé models in Equations (2.1)-(2.3). When the Scheffé specifications are directly embedded in a MNL framework, not all parameters are identified. To see this, we start with the special cubic model in Equation (2.3) and write the utility of alternative j in choice set s as

$$\begin{aligned} u_{js} &= \sum_{i=1}^{q} \beta_i x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^{q} \beta_{ik} x_{ijs} x_{kjs} + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^{q} \beta_{ikl} x_{ijs} x_{kjs} x_{ljs} + \varepsilon_{js} \\ &= \sum_{i=1}^{q-1} \beta_i x_{ijs} + \beta_q (1 - x_{1js} - \cdots - x_{q-1,js}) + \sum_{i=1}^{q-1} \sum_{k=i+1}^{q} \beta_{ik} x_{ijs} x_{kjs} + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^{q} \beta_{ikl} x_{ijs} x_{kjs} x_{ljs} + \varepsilon_{js} \\ &= \beta_q + \sum_{i=1}^{q-1} (\beta_i - \beta_q) x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^{q} \beta_{ik} x_{ijs} x_{kjs} + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^{q} \beta_{ikl} x_{ijs} x_{kjs} x_{ljs} + \varepsilon_{js}, \end{aligned}$$

where x_{ijs} denotes the proportion i of alternative j in choice set s. The second equality follows from the mixture constraint. The derivation demonstrates that a constant term is hidden in the q proportions x_i. As individuals choose the alternative with the highest utility, only utility differences are relevant in the MNL model (Ben-Akiva and Lerman, 1985; Franses and Paap, 2001). The constant in the utility specification therefore does not affect decisions and, hence, cannot be identified. As a result, we need to drop the (hidden) constant term.

The identified utility specification for the jth alternative in choice set s becomes

$$u_{js} = \sum_{i=1}^{q-1} \beta_i^* x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^{q} \beta_{ik} x_{ijs} x_{kjs} + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^{q} \beta_{ikl} x_{ijs} x_{kjs} x_{ljs} + \varepsilon_{js}, \quad (2.4)$$

where $\beta_i^* = \beta_i - \beta_q$. This notation highlights the fact that we should interpret β_i^* as the impact of proportion i relative to proportion q. We denote the vector of parameters in the identified model by $\boldsymbol{\beta}$.

2.3.3 Design optimality criterion

To construct an optimal experimental design, we use the D-optimality criterion. This is the most commonly used criterion in the literature on the design of mixture and choice experiments. Designs that are optimal with respect to this criterion in general also perform well in terms of other criteria, like the G-, V- and A-optimality criteria (Goos, 2002). A D-optimal design maximizes the determinant of the information matrix, or minimizes its inverse, the determinant of the variance matrix of the parameter estimator.

For the MNL model, the total information matrix $I(X, \boldsymbol{\beta})$ is obtained as the sum of the information matrices of the S choice sets, $I_s(X_s, \boldsymbol{\beta})$, and given by

$$\begin{aligned} I(X, \boldsymbol{\beta}) &= \sum_{s=1}^{S} I_s(X_s, \boldsymbol{\beta}) = \sum_{s=1}^{S} -\mathbb{E}(H_s(X_s, \boldsymbol{\beta})) \\ &= \sum_{s=1}^{S} -\mathbb{E}(-X_s'(P_s - \mathbf{p}_s\mathbf{p}_s')X_s) = \sum_{s=1}^{S} X_s'(P_s - \mathbf{p}_s\mathbf{p}_s')X_s, \end{aligned} \quad (2.5)$$

where $H_s(X, \boldsymbol{\beta})$ is the Hessian matrix of the log likelihood contribution of choice set s, $X_s = (\mathbf{f}(\mathbf{x}_{1s}), \mathbf{f}(\mathbf{x}_{2s}), \ldots, \mathbf{f}(\mathbf{x}_{Js}))'$, $X = (X_1', X_2', \ldots, X_S')'$, $\mathbf{p}_s = (p_{1s}, p_{2s}, \ldots, p_{Js})'$ and $P_s = \text{diag}(p_{1s}, p_{2s}, \ldots, p_{Js})$ (Ben-Akiva and Lerman, 1985). Note that, for the MNL model, the Hessian matrix does not depend on the response y and therefore the expectation operator can simply be dropped.

It is clear that the information matrix depends on the choice probabilities p_{js} and, hence, on the unknown parameter vector $\boldsymbol{\beta}$. In order to find a D-optimal design, we therefore need to assume (prior) values for the model parameters.

In the optimal experimental design literature, a *prior distribution* is often specified for the unknown parameters and a design is sought that minimizes the determinant of the inverse of the information matrix averaged over that prior distribution. The resulting design is referred to as the Bayesian optimal design (Atkinson et al., 2007). We denote the prior distribution for the parameter vector $\boldsymbol{\beta}$ by $\pi(\boldsymbol{\beta})$ and use the following Bayesian D-optimality criterion for mixture choice experiments for estimating the MNL model:

$$\mathrm{D} = \log\Big(\int_{\mathbf{R}^p} \{\det(I^{-1}(X, \boldsymbol{\beta}))\}^{1/p} \pi(\boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta}\Big), \tag{2.6}$$

where p is the number of parameters in $\boldsymbol{\beta}$. The exponent $1/p$ makes the value of the D-criterion comparable across models with different numbers of parameters. Note that the log transformation has no impact on the resulting optimal design. The Bayesian D-optimality criterion in Equation (2.6) was introduced in the context of choice experiments by Sándor and Wedel (2001) and it has been used by many other authors studying the D-optimal design of choice experiments. Our criterion corresponds to Criterion II in Atkinson et al. (2007) and is appropriate when the variance of the parameter estimates is most important.

Through the prior distribution, we can express different prior beliefs about the parameters. It is possible to use a degenerate prior distribution putting all weight on a single value for the parameter vector, say $\tilde{\boldsymbol{\beta}}$. The Bayesian D-optimality criterion then simplifies to the local D-optimal design criterion given by $\log(\det(I^{-1}(X, \tilde{\boldsymbol{\beta}}))^{1/p})$. When $\tilde{\boldsymbol{\beta}} = \mathbf{0}_p$, the locally D-optimal design is called the *D-optimal utility-neutral design*.

Specifying a prior distribution for the parameters in the identified MNL model, involving the differences $\beta_1 - \beta_q, \ldots, \beta_{q-1} - \beta_q$, is not straightforward as we would like the prior information to be independent of the chosen baseline ingredient. We recommend first specifying a normal prior distribution for the parameters of the full (unidentified) Scheffé model (as in Equations (2.1)-(2.3)), for example,

$$\mathcal{N}(\mathbf{0}, \kappa I_{p+1}), \tag{2.7}$$

where κ is a scalar measuring prior uncertainty, and next transforming it into a prior for the identified model parameters (see Equation (2.4)). As the required transformation is linear, the resulting prior distribution $\pi(\boldsymbol{\beta})$ for the identified model is still multivariate normal.

However, due to the transformation, the parameters are no longer independent a priori. The resulting prior specification is invariant to the reference proportion chosen.

In order to construct a D-optimal experimental design, a numerical evaluation of the p-dimensional integral in Equation (2.6) with respect to the prior distribution $\pi(\boldsymbol{\beta})$ is required. For this numerical evaluation, we use a simulation-based method, where we sample parameter values from the prior distribution. The computation time scales linearly with respect to the number of draws. To shorten computing times, we use a systematic sample from the prior distribution based on Halton sequences (Halton, 1960). Spanier and Maize (1994) demonstrated that a small number of Halton draws provide a relatively good integration. In the context of choice experiments, Bliemer et al. (2008) argue that using Halton sampling results in more efficient designs than using pseudo Monte Carlo draws. For further advice, we refer to Yu et al. (2010). In order to speed up the computations even further, we use the Cholesky decomposition to calculate the determinant of the information matrix.

2.4 Design construction algorithms

In this section, we present two algorithms for the construction of optimal choice designs involving mixtures, namely, a particle swarm optimization and a mixture coordinate-exchange algorithm. Both methods minimize the D-optimality criterion through a local search procedure and involve two main steps: (1) the generation of a starting design and (2) the improvement of the starting design. The algorithms are heuristic optimization methods and, therefore, there is no guarantee that they will produce the globally optimal design. To speed up convergence and to increase the likelihood of finding the global optimum, we consider a number of smart, high quality starting designs. The mixture coordinate-exchange algorithm will be applied to each of these starting designs. In our particle swarm optimization algorithm, we use these starting designs as initial particles. The approach of considering many different starting designs is common in algorithms for constructing optimal experimental designs (Atkinson et al., 2007).

Below, we first describe how we obtain starting designs. We then turn to the particle swarm optimization algorithm and the mixture coordinate-exchange algorithm.

2.4.1 Generating a starting design

To improve the performance of the optimization methods, we use starting designs that are optimal for a simplified problem. More specifically, we use the method introduced by Goos and Donev (2007). We choose the p distinct mixtures which have been proven to be D-optimal for the linear (first-order, second-order or special-cubic) Scheffé model under consideration in case a completely randomized experiment would be performed. Next, we replicate these p distinct mixtures as evenly as possible, to obtain the required number of mixtures ($n = SJ$). Finally, we randomly allocate these n mixtures to the S choice sets, subject to the constraint that a given mixture should not appear more than once in any choice set. The resulting design is now D-optimal in the class of minimum support designs for the Scheffé model under consideration in case there are S blocks of size J. A minimum support design is a design which has as many distinct mixtures as there are parameters in the model. For more details, we refer to Goos and Donev (2007).

Due to the equivalence of the problem of finding a D-optimal design for a blocked experiment for a linear model and the D-optimal choice design problem in case the parameters are zero (Kessels et al., 2011b), the resulting starting designs are D-optimal for a problem that approximates the Bayesian D-optimal mixture choice design problem. Therefore, the starting design is likely to be of high quality. Furthermore, this procedure for generating starting designs involves many random components. It is therefore capable of producing a large number of different high quality starting designs. Our computational experiments showed that using the minimum support designs rather than completely random starting designs substantially decreases the required computation time of the mixture coordinate-exchange algorithm. For the particle swarm optimization algorithm, the minimum support starting design does not speed up the computations.

2.4.2 Improving the starting designs

Particle swarm optimization

Particle swarm optimization (PSO) was originally developed by Kennedy and Eberhart (1995) and Shi and Eberhart (1998). It is inspired by the social behavior of bird flocking or fish schooling. The PSO algorithm operates by considering a randomly generated set (the so-

called swarm) of candidate designs (the so-called particles) and sequentially updating the particles towards the optimum. Each particle's update is governed by a set of equations that describe its new location as a function of its current location and its velocity. When moving towards the optimum, the particles share the knowledge that they have accumulated concerning the location of the optimum.

We employ the approach of Wong et al. (2015) who redefine their mixture design problem as an optimization problem on a hypercube instead of on a simplex by introducing a projection function that maps the hypercube to the simplex. This simplifies and speeds up the optimization in the context of the PSO algorithm. An alternative i in a design is represented by a vector $\mathbf{z}_i$ in the q-dimensional unit hypercube, excluding the vector $\mathbf{0}$. We denote this space by Ξ. A complete design is represented by a point $\tilde{\boldsymbol{\xi}} = (\mathbf{z}_1', \mathbf{z}_2', \ldots, \mathbf{z}_n')'$ in Ξ^n. Proper mixture alternatives are obtained by rescaling each $\mathbf{z}_i$ such that the sum of its elements equals 1. Applying this transformation to each vector $\mathbf{z}_i$ in $\tilde{\boldsymbol{\xi}}$ yields a valid design $\boldsymbol{\xi}$. We denote this projection as $\boldsymbol{\xi} = P(\tilde{\boldsymbol{\xi}})$.

The PSO procedure can now be described as follows. First, initialize a population of N candidate designs $\tilde{\boldsymbol{\xi}}_j^0$ from Ξ^n, the so-called particles. During the algorithm, we keep track of (1) $\tilde{\boldsymbol{\xi}}_j^{pbest}$, the best position that the j^{th} particle has ever visited; and (2) $\tilde{\boldsymbol{\xi}}^{gbest}$, the best position taken by any of the particles during the entire execution of the algorithm. Then, at iteration t, we update the location of the j^{th} particle using $\tilde{\boldsymbol{\xi}}_j^t = \tilde{\boldsymbol{\xi}}_j^{t-1} + \chi \mathbf{v}_j^t$, where χ is a pre-specified positive constant and $\mathbf{v}_j^t$ is the so-called velocity with as k^{th} element v_{kj}^t. If $\tilde{\boldsymbol{\xi}}_j^t$ is not in Ξ^n, it is projected to the closest point on the boundary of Ξ^n. This means that all elements above 1 are mapped to 1 and all elements below 0 are mapped to 0. The velocity determines the direction in which the particle moves. Denote the k^{th} element, $k = 1, \ldots, nq$, of the j^{th} particle $\tilde{\boldsymbol{\xi}}_j^t$ at iteration t by $\tilde{\xi}_{kj}^t$ and the k^{th} elements, $k = 1, \ldots, nq$, of $\tilde{\boldsymbol{\xi}}_j^{pbest}$ and $\tilde{\boldsymbol{\xi}}^{gbest}$ by $\tilde{\xi}_{kj}^{pbest}$ and $\tilde{\xi}_k^{gbest}$, respectively. For element k of particle $\tilde{\boldsymbol{\xi}}_j^t$, the velocity at iteration t is defined as

$$v_{kj}^t = w_t v_{kj}^{t-1} + c_1 \epsilon_k^1 (\tilde{\xi}_k^{gbest} - \tilde{\xi}_{kj}^{t-1}) + c_2 \epsilon_k^2 (\tilde{\xi}_{kj}^{pbest} - \tilde{\xi}_{kj}^{t-1}), \tag{2.8}$$

where w_t is the inertia weight, c_1 and c_2 are pre-specified positive constants and ϵ_k^1 and ϵ_k^2 are uniform$(0, 1)$ random numbers. A maximum value of one is used for each particle's velocity. Next, we calculate the D-value for particle $\tilde{\boldsymbol{\xi}}_j^t$ at iteration t using the design $\boldsymbol{\xi}_j^t = P(\tilde{\boldsymbol{\xi}}_j^t)$ and

update $\tilde{\boldsymbol{\xi}}_j^{pbest}$ and $\tilde{\boldsymbol{\xi}}^{gbest}$ if necessary. We iterate this procedure over j and t until convergence or until a maximum number of iterations has been reached. Finally, we obtain the optimal design as $\boldsymbol{\xi}^{gbest} = P(\tilde{\boldsymbol{\xi}}^{gbest})$. As initial velocities, we take $v_{kj}^0 = 0$. We provide pseudo code for our PSO algorithm in Appendix 2.C.

The PSO algorithm has a number of tuning parameters. The constants c_1 and c_2 are the cognitive and the social learning factors, respectively. Following Kennedy (1997) and Wong et al. (2015), we set both factors to 2. The constant χ is the constriction factor, which is usually set to 1. Following Wong et al. (2015), we let w_t decrease linearly from 0.9 to 0.4 over the maximum number of iterations. Our experimental results are in line with those of Wong et al. (2015) in that the optimal designs seem insensitive to different settings of tuning parameters.

In order to use the starting designs discussed in Section 2.4.1 as initial particles, we have to project their alternatives from the $(q-1)$-simplex to $\boldsymbol{\Xi}$. However, this mapping is one-to-many. We therefore introduce a random factor in this mapping. For a specific alternative, we multiply all proportions by a draw from the uniform$(0, \frac{1}{\max \text{Proportion}})$ distribution, where $\max \text{Proportion}$ is the largest proportion in the original mixture. The resulting point $\mathbf{z}_i$ is guaranteed to be an element of $\boldsymbol{\Xi}^n$ and still maps back onto the original starting design.

Mixture coordinate-exchange algorithm

The second optimization method we consider is a mixture coordinate-exchange algorithm (Piepel et al., 2005; Meyer and Nachtsheim, 1995). Given an initial design, the algorithm starts by optimizing the first ingredient proportion of the first alternative in the first choice set using the method of Brent (1973). Brent's method is a one-dimensional optimization algorithm based on a combination of golden section search and successive parabolic interpolations. The mixture coordinate-exchange algorithm then continues by optimizing the second ingredient proportion of the first alternative in the first choice set, the third ingredient proportion, ..., and the qth ingredient proportion. This process is repeated for each alternative in each choice set of the design. Whenever at least one proportion has been improved in a complete pass of the coordinate-exchange algorithm through the entire design, the optimization procedure is restarted from the first ingredient proportion. The algorithm stops when no improvements have been performed in a complete pass through all the proportions.

This mixture coordinate-exchange algorithm differs in a substantial way from the original coordinate-exchange algorithm in Meyer and Nachtsheim (1995). Because proportions cannot be changed independently, we adopt the approach by Piepel et al. (2005): when optimizing a proportion x_i, we keep the pairwise ratios of the other ingredient proportions fixed. This approach is known as exchanges along Cox-effect directions (Piepel, 1982; Cornell, 2002). The general expression for recomputing the proportions $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_q$ after a proportion x_i is changed to $x_i + \Delta$ is

$$x_j^{\text{new}} = x_j \left(1 - \frac{\Delta}{1 - x_i}\right).$$

In case $x_i = 1$ before the change, we set the proportions $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_q$ to $\Delta/(q-1)$. We provide pseudo code of our mixture coordinate-exchange algorithm in Appendix 2.D.

Our algorithmic approach differs from the mixture coordinate-exchange algorithm of Piepel et al. (2005) for linear Scheffé models and completely randomized experiments in two important ways. First, we use a procedure for generating a high quality starting design, which takes into account the fact that a design for a mixture choice experiment requires the mixtures to appear in groups. Second, we have adapted the algorithm so that it optimizes the Bayesian D-optimality criterion for the MNL model.

2.5 Design performance and illustrations

In this section, we illustrate our approach and compare the performance of different designs and the two optimization algorithms. First, we consider utility-neutral designs, i.e., we set $\boldsymbol{\beta} = \mathbf{0}_p$. For this scenario, we provide a thorough comparison of the PSO and the mixture coordinate-exchange algorithms and we show that the mixture coordinate-exchange algorithm provides more efficient designs. When running the PSO algorithm many times, it consistently obtains higher (worse) D-values. For the detailed design comparisons in the subsequent applications, we therefore rely on the mixture coordinate-exchange algorithm only. However, to demonstrate the good performance of the mixture coordinate-exchange algorithm in all circumstances, we consistently present the designs and the D-values obtained by the PSO algorithm for all scenarios we consider. We also demonstrate that, for the utility-neutral designs, the performance of our mixture coordinate-exchange algorithm matches that

of the mixture coordinate-exchange algorithm of Piepel et al. (2005), as implemented in the software package JMP10. The excellent performance of our algorithm in this case provides confidence that the algorithm is also capable of finding good designs for other, more complex, cases that cannot be tackled using standard software. Next, we consider locally optimal designs for fixed, non-zero, values of $\boldsymbol{\beta}$ and show that these designs outperform the utility-neutral design when both designs are evaluated at $\boldsymbol{\beta}$. Finally, we study Bayesian D-optimal designs.

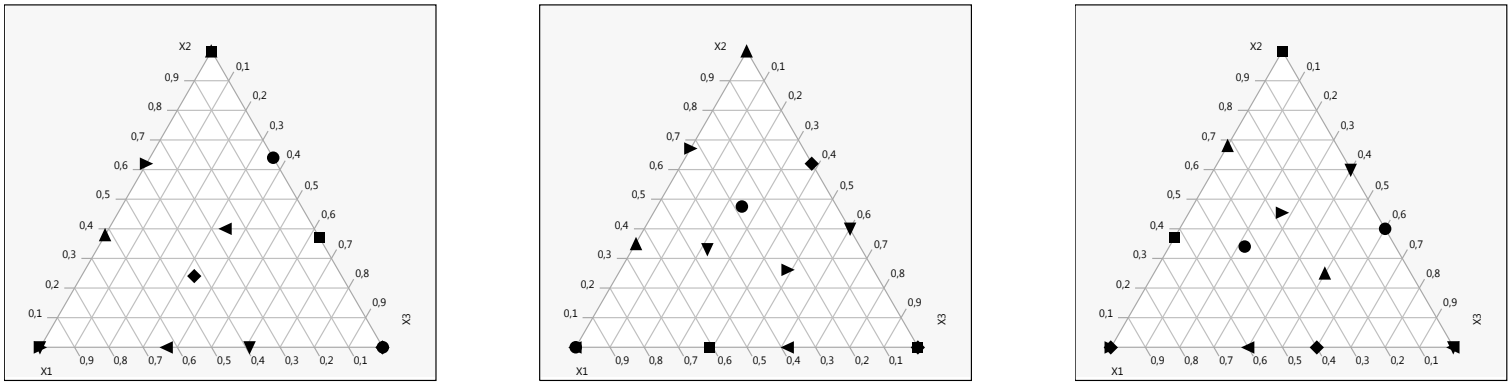
Although both the PSO algorithm and the mixture coordinate-exchange algorithm are capable of handling any number of ingredients and any kind of Scheffé model, in this section, for illustrative purposes, we only report designs for the special-cubic model for three ingredients with 14 alternatives in seven choice sets of size two. All reported designs are obtained using $1,000$ different starting designs for the mixture coordinate-exchange algorithm and $1,000$ particles for the PSO algorithm. We choose the same number of starting designs and particles, in order to compare both methods on a fair basis. For the Bayesian D-optimal designs, we used 128 Halton draws to approximate the Bayesian D-optimality criterion. The computations were performed using Matlab on a machine with an AMD Athlon II X2 B28 (3.40 GHz) processor and 4 GB RAM.

2.5.1 Utility-neutral designs

We first construct choice designs assuming that $\boldsymbol{\beta} = \mathbf{0}_p$. We do so for two reasons. First, as pointed out in Section 2.2, several authors in the choice design literature focus on this scenario. Second, studying the case where $\boldsymbol{\beta} = \mathbf{0}_p$ allows us to compare the performance of our algorithm to that of commercial software, as the problem of finding an optimal design for a blocked experiment is equivalent to that of finding an optimal choice design under the assumption of $\boldsymbol{\beta} = \mathbf{0}_p$.

In Figure 2.1, we show the utility-neutral designs generated by the PSO algorithm, our mixture coordinate-exchange algorithm and JMP10 using ternary plots. We provide the ingredient proportions for each of the designs in tabular form in Appendix 2.A in Tables A.1, A.2 and A.3, respectively. The vertices of the triangles in Figure 2.1 correspond to the three alternatives where one proportion equals 100%. Interior points represent mixtures in which none of the three ingredients is absent, i.e., where x_1, x_2 and x_3 are all strictly positive. The

centroid corresponds to the mixture with equal proportions for each of the three ingredients (Cornell, 2002). Each symbol in the ternary plots corresponds to a different choice set. As certain mixtures in Figure 2.1 appear in more than one choice set, some symbols overlap. In total, there are seven symbols in each ternary plot, each of which occurs twice, as we consider designs with seven choice sets of two alternatives.



(a) PSO algorithm **(b)** Mixture coordinate-exchange algorithm **(c)** JMP10

Figure 2.1: Utility-neutral designs

The design obtained using the PSO algorithm looks quite different from the designs obtained using the mixture coordinate-exchange algorithm and JMP10. The locations and grouping of the mixtures in the two latter designs are remarkably similar (Figures 2.1b and 2.1c). The D-optimality criterion values equal 2.9468 for the PSO algorithm, 2.9397 for the design constructed using our mixture coordinate-exchange algorithm, and 2.9410 for the design constructed using JMP10. Thus, the design produced by the mixture coordinate-exchange algorithm performs slightly better than the one obtained using JMP10. The design constructed using the PSO algorithm performs worst. However, the running time of the PSO algorithm was much smaller (369 seconds CPU time) than that for our mixture coordinate-exchange algorithm ($1,092$ seconds). To give the PSO algorithm an equal chance, we ran it three times, resulting in a total CPU time of $1,094$ seconds. However, this did not result in a better design.

Especially our mixture coordinate-exchange algorithm works well in this scenario. Evaluating the performance of the mixture coordinate-exchange algorithm in several other utility-neutral scenarios yields similar results: the PSO algorithm never produced a lower D-value

than our mixture-coordinate exchange algorithm. Therefore, in the next sections, we refer to our mixture coordinate-exchange algorithm as the main algorithm.

2.5.2 Locally optimal designs

Constructing locally optimal designs

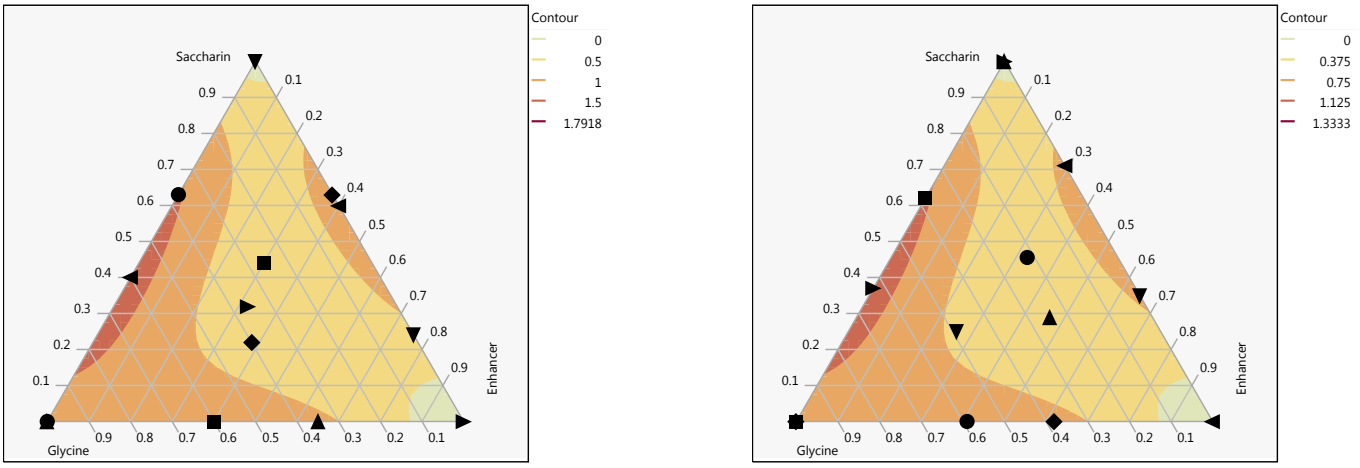
Using parameter values equal to zero may seem attractive; it is convenient and helps to reduce the computation time required to generate an optimal design, if compared to Bayesian or locally optimal designs. However, the corresponding preference assumption is unrealistic, since it is hard to believe that respondents are indifferent across all alternatives. For instance, consumers generally prefer low prices over high ones and certain brands are consistently preferred over others. In a mixture context, not all ingredients might be equally important to the respondents.

In this section, we demonstrate how we can use consumer preference information to construct locally optimal experimental designs assuming a particular value for $\boldsymbol{\beta}$. In order to use realistic parameter values, we analyze data from a three-ingredient mixture experiment in Cornell (2002) meant to investigate whether an artificial sweetener could be used in a popular athletic-sports drink. The three ingredients considered were glycine (x_1), saccharin (x_2) and an enhancer (x_3). The response of interest was "intensity of sweetness aftertaste". The special-cubic model estimated based on rank-order data from 380 individuals is given by

$$\hat{y} = 11.25x_1 + 5.54x_2 + 3.73x_3 + 26.93x_1x_2 + 20.52x_1x_3 + 28.44x_2x_3 - 180.68x_1x_2x_3.$$

We interpret the intensity of sweetness aftertaste as being proportional to the utility in the MNL model. Now, because we embed the special-cubic model in the MNL model, we need to account for identification restrictions and, therefore, the corresponding utility model becomes

$$\begin{aligned}\hat{u} &= (11.25 - 3.73)x_1 + (5.54 - 3.73)x_2 + 26.93x_1x_2 + 20.52x_1x_3 + 28.44x_2x_3 - 180.68x_1x_2x_3 \\ &= 7.52x_1 + 1.81x_2 + 26.93x_1x_2 + 20.52x_1x_3 + 28.44x_2x_3 - 180.68x_1x_2x_3.\end{aligned}$$



(a) Odds of 6 **(b)** Odds of 3.8

Figure 2.2: Locally optimal experimental designs produced by the mixture coordinate-exchange algorithm for the rescaled parameter vectors $\boldsymbol{\beta}_1$ (odds of 6) and $\boldsymbol{\beta}_2$ (odds of 3.8) in the intensity of sweetness aftertaste example

For these parameter values, the utility ranges from 0 to 11.70, which results in an extreme odds ratio of $\exp(11.70)/\exp(0) > 1.2 \times 10^5$ for choosing the best available alternative over the worst, which is clearly unrealistic. Therefore, prior to constructing locally optimal designs, we first rescale the parameter values such that the maximum odds ratio decreases to 6. Next, we further rescale the parameters to make the maximum odds ratio decrease to 3.8. In doing so, we make sure not to change the shape of the utility function. In these two new scenarios, the utility values are in the intervals $[0, 1.79]$ and $[0, 1.33]$, respectively. Note that, when the utility range decreases, we approach the utility-neutral case, for which the utility range is $[0, 0]$. The rescaled parameters are $\boldsymbol{\beta}_1 = (1.15, 0.28, 4.12, 3.14, 4.36, -27.67)'$ and $\boldsymbol{\beta}_2 = (0.86, 0.21, 3.07, 2.34, 3.24, -20.59)'$.

We graphically show the locally optimal designs produced by the mixture coordinate-exchange algorithm for the two scenarios in Figure 2.2 and again provide the designs in tabular format in Appendix 2.A (Tables A.4 and A.5 for the coordinate-exchange algorithm and Tables A.6 and A.7 for the PSO algorithm). The shaded areas in Figure 2.2 give the contour lines of the utility function. The D-values of the two designs produced by the PSO algorithm are 3.1103 and 3.0246, respectively, while the D-values produced by the mixture coordinate-exchange algorithm are 3.0474 and 3.0132.

The two designs in Figure 2.2a and 2.2b differ substantially, showing that the a priori information provided has a considerable impact on the design. As expected, the design

constructed for the narrowest utility range, i.e., the design in Figure 2.2b, is most similar to the utility-neutral design in Figure 2.1b.

Comparing locally optimal to utility-neutral designs

When assessing the performance of a utility-neutral design in case $\boldsymbol{\beta} \neq \mathbf{0}_p$, we need to take into account the fact that the optimal solution to the utility-neutral design problem is not unique. If $\boldsymbol{\beta} = \mathbf{0}_p$, all column permutations of any utility-neutral design matrix result in a design with the same D-value as the original. Therefore, it is possible to construct $q!$ different, but equivalent designs by permuting the labels of all ingredients when $\boldsymbol{\beta} = \mathbf{0}_p$. In Appendix 2.B, we list all these permutations of the utility-neutral design in Figure 2.1b.

The permuted designs are not equivalent when evaluated under a non-zero parameter vector. As the choice of the permutation to eventually use in an experiment is arbitrary, we consider all permutations when comparing the locally optimal design to the utility-neutral approach. For each non-zero parameter vector we considered above ($\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$), we identify the best permutation of the utility-neutral design as well as the worst permutation of that design. The D-optimality criterion values for the best and worst utility-neutral designs are 3.0704 and 3.1284, respectively, for $\boldsymbol{\beta}_1$ and 3.0147 and 3.0482, respectively, for $\boldsymbol{\beta}_2$. Recall that the D-values for the locally optimal designs produced by the mixture coordinate-exchange algorithm are 3.0474 for $\boldsymbol{\beta}_1$ and 3.0132 for $\boldsymbol{\beta}_2$.

Although the locally optimal designs perform better in both scenarios, the difference in performance decreases as the prior parameter vector approaches the zero vector. For parameter vector $\boldsymbol{\beta}_1$, the locally optimal design is 0.33% more efficient than the best and 1.16% more efficient than the worst utility-neutral design. The difference in efficiency drops to 0.02% and 0.5% for the best and worst utility-neutral designs, respectively, for $\boldsymbol{\beta}_2$. Therefore, our results show that the conclusion of Huber and Zwerina (1996) that non-zero priors can be used to generate statistically more efficient choice designs extends from choice experiments involving categorical attributes to choice experiments involving mixtures.

Another common statement in the literature is that utility balance is a good property for an experimental design. Both our locally optimal designs turn out to be approximately utility balanced. This means that the alternatives within any given choice set tend to have similar utilities and, hence, similar choice probabilities. This is not always the case for the utility-

neutral designs. For a perfectly utility-balanced design with choice sets of two alternatives, the product of the two choice probabilities in any choice set should be 0.25. If the alternatives differ in utility, this product is smaller than 0.25.

Consider again the two locally optimal designs, constructed assuming $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_2$, and the corresponding best and worst utility-neutral designs. For each of these designs, we provide the minimum, mean and maximum values of the product of the choice probabilities across choice sets in Table 2.1. The table shows that the locally optimal design approach always scores very well in terms of utility balance, whereas the utility-neutral designs exhibit a lack of utility balance. Under $\boldsymbol{\beta}_1$, the utility-neutral designs have choice sets for which the choice probabilities' product equals 0.15, in which case the choice probabilities are as unequal as 0.2 and 0.8. None of the two locally optimal designs results in choice probabilities that differ to this extent.

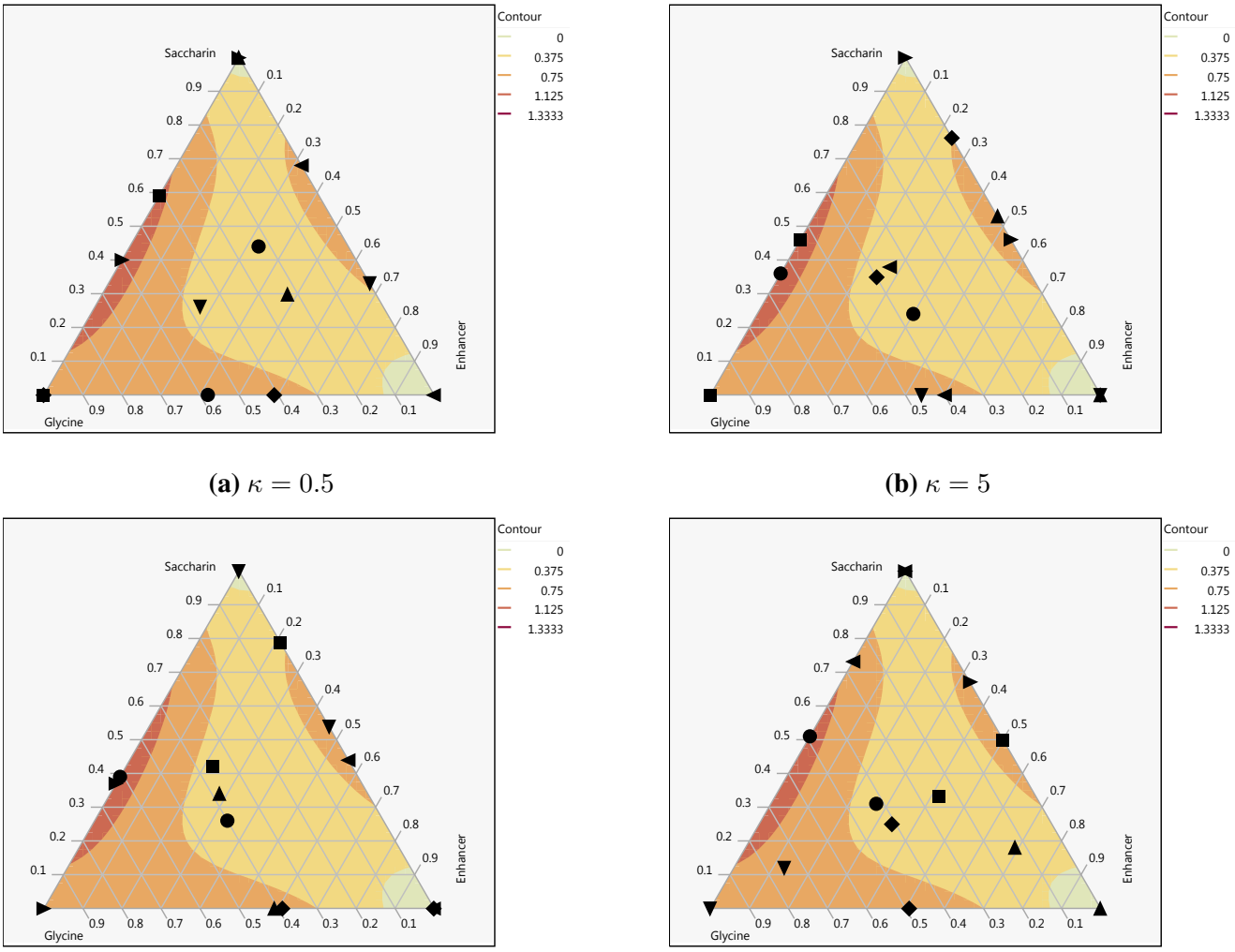
	Minimum			Mean			Maximum		
	Locally optimal	Best UN	Worst UN	Locally optimal	Best UN	Worst UN	Locally optimal	Best UN	Worst UN
$\boldsymbol{\beta}_1$	0.22	0.15	0.15	0.23	0.22	0.21	0.25	0.25	0.25
$\boldsymbol{\beta}_2$	0.18	0.18	0.19	0.23	0.23	0.22	0.25	0.25	0.25

Table 2.1: Measures of utility balance for the locally optimal and utility-neutral (UN) designs

2.5.3 Bayesian optimal designs

Bayesian D-optimal designs, constructed using a proper prior distribution that incorporates a priori uncertainty about the model parameters, are known to perform well under a broad range of scenarios (Sándor and Wedel, 2001; Kessels et al., 2011b). In this section, we investigate the impact of accounting for this uncertainty using independent prior distributions for the (unidentified) parameters. We show how to deal with a multivariate prior distribution involving correlations between the different parameters in the next section using the example of Courcoux and Séménou (1997).

As a prior distribution, we choose a normal distribution with prior mean equal to $\boldsymbol{\beta}_2$ and variance matrix κI_{p+1} (see Equation (2.7)) transformed to the identified parameter space. The a priori uncertainty is controlled by the parameter κ, for which we use the values 0.5, 5, 10 and 30. We show the Bayesian D-optimal designs corresponding to the different κ values in Figure 2.3 and provide their design points in tabular format in Appendix 2.A (Tables A.8



(a) $\kappa = 0.5$ **(b)** $\kappa = 5$

(c) $\kappa = 10$ **(d)** $\kappa = 30$

Figure 2.3: Bayesian optimal designs for $\boldsymbol{\beta}_2$ and different levels of uncertainty

$-$ A.11 for the mixture coordinate-exchange algorithm and Tables A.12 $-$ A.15 for the PSO algorithm). The contours in Figure 2.3 correspond to the mean parameter vector ($\boldsymbol{\beta}_2$). When the uncertainty increases, the design becomes less symmetric and the design points tend to be spread across the entire simplex. We also find that the design for $\kappa = 0.5$ is almost identical to the locally optimal design in Figure 2.2b, which implicitly assumes $\kappa = 0$. The D-values for the designs obtained by the mixture coordinate-exchange algorithm are 3.0769, 3.4904, 3.7698 and 4.3759, respectively. The corresponding D-values for the PSO algorithm are 3.1076, 3.5178, 3.7752 and 4.4522, respectively.

A remarkable feature of the Bayesian D-optimal designs is that, for larger values of κ, the mixtures within a choice set tend to be close to each other. To quantify this phenomenon, we computed the minimum, mean and maximum Euclidean distances between the two alter-

natives within a choice set for every design. The results are shown in Table 2.2. Clearly, the distances decrease substantially when the prior uncertainty, as measured by κ, is increased. The distances for the Bayesian D-optimal design when $\kappa = 30$ are only half as large as for the locally optimal design, for which $\kappa = 0$.

	Locally Optimal	Bayesian D-Optimal			
	$\kappa = 0$	$\kappa = 0.5$	$\kappa = 5$	$\kappa = 10$	$\kappa = 30$
min	0.59	0.58	0.51	0.42	0.31
mean	0.82	0.79	0.63	0.53	0.36
max	1.00	0.96	0.76	0.65	0.46

Table 2.2: Euclidean distances between alternatives within a choice set for the locally optimal design and its Bayesian counterparts

To show that the Bayesian D-optimal designs also exhibit approximate utility balance, we provide the minimum, mean and maximum values of the product of choice probabilities in a choice set in Table 2.3, along with the corresponding values for the locally optimal design constructed at $\boldsymbol{\beta}_2$ (for which $\kappa = 0$). The mean values are all very close to 0.25, while the maximum values are 0.25 for all the designs. Because the locally optimal design and the Bayesian D-optimal design for $\kappa = 0.5$ are very similar, the utility-balance measures for the two designs are identical.

	Locally Optimal	Bayesian D-Optimal			
	$\kappa = 0$	$\kappa = 0.5$	$\kappa = 5$	$\kappa = 10$	$\kappa = 30$
min	0.18	0.18	0.20	0.21	0.22
mean	0.23	0.23	0.22	0.23	0.23
max	0.25	0.25	0.25	0.25	0.25

Table 2.3: Utility balance measures for the locally optimal design and its Bayesian counterparts

2.6 Quantifying cocktail preferences

As a final illustration, we reconsider the experiment for quantifying cocktail preferences as described in Courcoux and Séménou (1997). The authors consider seven fruit cocktails that involve three ingredients: mango juice (x_1), blackcurrant syrup (x_2) and lemon juice (x_3), which we plot in Figure 2.5a. In the experiment, a panel of 60 consumers was asked to taste different pairs of the seven fruit cocktails and to indicate the preferred cocktail in each pair. Each respondent had to evaluate eight of the 21 possible pairs. As a result, the final experimental design contained $60 \times 8 = 480$ choice sets of size two in total.

In this section, we compare designs built by using our approach to the ad-hoc design used by Courcoux and Séménou (1997). First, since, before conducting the experiment, one may not have prior information about the parameter values, we obtain a Bayesian D-optimal design using an uninformative prior as in Equation (2.7) with $\kappa = 5$ and assuming the special-cubic Scheffé model. As obtaining a Bayesian D-optimal experimental design with 480 choice sets of size two would be computationally very demanding, we construct an experimental design with 32 alternatives in 16 choice sets of size two. Since Courcoux and Séménou (1997) imposed lower bounds of 0.3, 0.15 and 0.1 on the three ingredient proportions, we redefine the coordinates of the simplex in terms of pseudo components (see Section 2.3.1). The optimal designs in terms of the pseudo components are tabulated in Tables A.16 (coordinate-exchange) and A.17 (PSO) in Appendix 2.A. The Bayesian D-values of the two designs are 2.5233 for the coordinate-exchange algorithm and 2.5592 for the PSO algorithm.

To compare our Bayesian D-optimal experimental design to the original design used by Courcoux and Séménou (1997), we compare the D-values that both designs yield for different potential true parameter values. In order to obtain a true preference distribution, we use a mixed logit model (Train, 2009) to estimate the parameter means and variances, starting from the choice probabilities reported in Courcoux and Séménou (1997) and assuming a special-cubic Scheffé model. The preference distribution which we obtain for the identified model (see the discussion in Section 2.3.3) is $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta_0}, \Sigma_0)$, with mean vector $\boldsymbol{\beta_0} = (1.36, 1.57, 2.47, -0.43, 0.50, 1.09)'$ and variance matrix

$$\Sigma_0 = \begin{pmatrix} 6.14 & 5.00 & 2.74 & -0.43 & -2.81 & -3.33 \\ 5.00 & 6.76 & 4.47 & -1.79 & -6.13 & -3.51 \\ 2.74 & 4.47 & 3.45 & -1.38 & -4.71 & -2.17 \\ -0.43 & -1.79 & -1.38 & 1.18 & 2.39 & 0.71 \\ -2.81 & -6.13 & -4.71 & 2.39 & 7.43 & 2.71 \\ -3.33 & -3.51 & -2.17 & 0.71 & 2.71 & 2.49 \end{pmatrix}.$$

Next, we compute D-values for $10,000$ random draws from this preference distribution for the original design used in Courcoux and Séménou (1997) and our Bayesian D-optimal experimental design obtained with an uninformative prior. Note that, in order to compare the

information matrix of the original design (which contains 480 choice sets of size two) to the information matrix of our design (which contains 16 choice sets of size two), we calculate the information matrix for our design as $I = 30 \sum_{s=1}^{16} I_s$. In doing so, we assume that our design is replicated 30 times, resulting in 480 choice sets as well. We provide the distributions of $10,000$ D-values for both designs in Figure 2.4. The means of the D-criterion values are equal to -0.3101 and -0.2859 for the original and Bayesian designs, respectively. Note that the negative D-values are due to the log transformation of determinant values smaller than 1 (see Equation (2.6)). Bayesian design performed better in 41.34% of the cases. As we can see, both designs perform remarkably similar, despite the fact that Courcoux and Séménou (1997) considered a heterogeneous choice design, which, according to Sándor and Wedel (2005), tends to be more efficient than a homogeneous choice design. This demonstrates that our approach is capable of generating experimental designs of high quality even when an uninformative prior distribution is used.

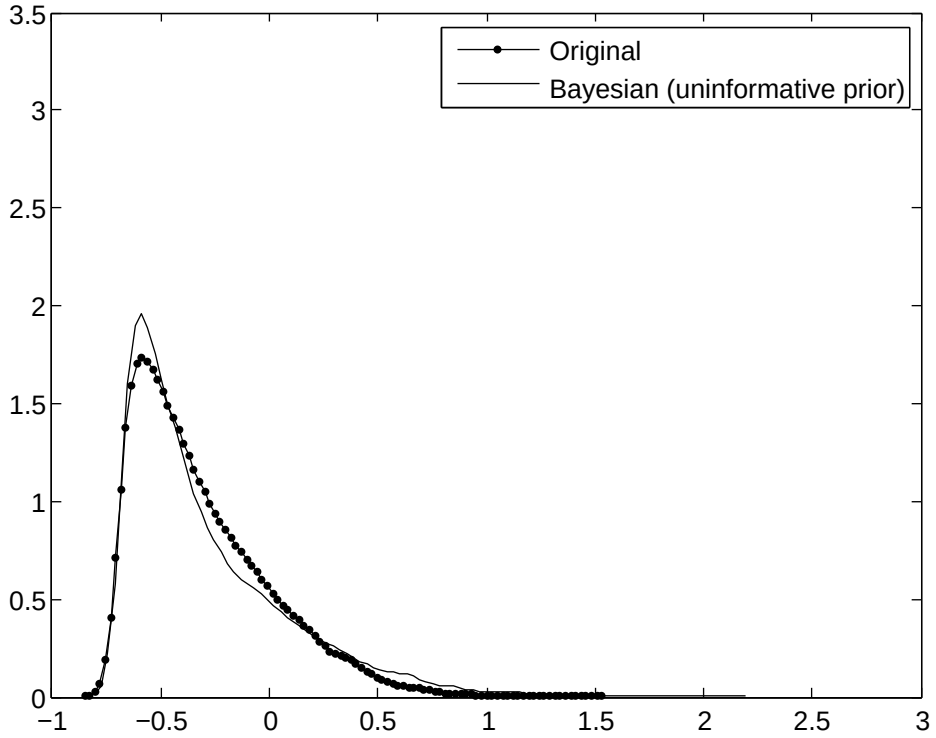
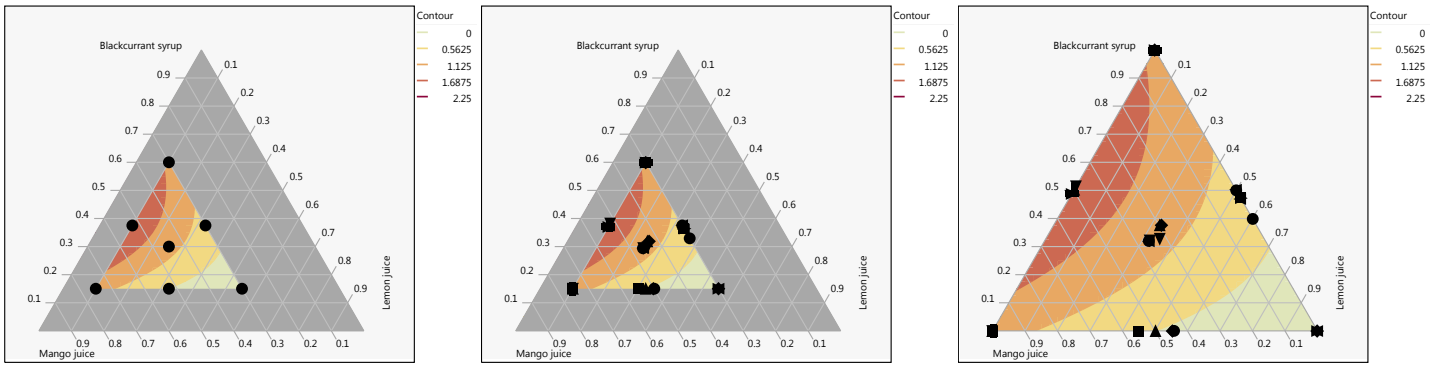


Figure 2.4: Distribution of D-optimality criterion values for the original design and Bayesian D-optimal design obtained with an uninformative prior

We next take the preference distribution $\mathcal{N}(\boldsymbol{\beta}_0, \Sigma_0)$ as a prior distribution for computing a Bayesian D-optimal design for the same experiment. The design obtained using the coordinate-exchange algorithm is presented in two ways in Figure 2.5, once in terms of the

true ingredients (Figure 2.5b) and once in terms of the pseudo components (Figure 2.5c). Figure 2.5 also shows the contours of the expected utility function corresponding to $\boldsymbol{\beta}_0$. The optimal designs in terms of the pseudo components are given in tabular format in Tables A.18 (coordinate-exchange) and A.19 (PSO) in Appendix 2.A. The Bayesian D-values of the two designs are 2.7153 for the coordinate-exchange algorithm and 2.7792 for the PSO algorithm.



(a) Mixtures used in Courcoux and Séménou (1997)

(b) Bayesian design in the original components

(c) Bayesian design in the pseudo components

Figure 2.5: The design points used by Courcoux and Séménou (1997) together with the Bayesian D-optimal design obtained using a prior of $\mathcal{N}(\boldsymbol{\beta}_0, \Sigma_0)$

We also obtain a locally optimal design for parameter vector $\boldsymbol{\beta_0}$ (assuming no uncertainty) and utility-neutral designs for this experiment. We provide the utility-neutral designs, expressed in terms of the pseudo components, produced by the coordinate-exchange and PSO algorithms in Tables A.20 and A.21, respectively, in Appendix 2.A. The corresponding D-values are 2.1500 and 2.1652. The utility-neutral designs produced by the mixture coordinate-exchange algorithm that result in the best and worst Bayesian D-values appear in Tables A.22 and A.23 of Appendix 2.A. The locally optimal design is tabulated in terms of the pseudo components in Table A.24 for the coordinate-exchange and in Table A.25 for the PSO algorithm. The D-value for the coordinate-exchange algorithm equals 2.2809, while it equals 2.3061 for the PSO algorithm.

We now study how the Bayesian D-optimal (Figures 2.5b-2.5c), locally optimal and the best and worst utility-neutral designs perform in comparison to the original design from Courcoux and Séménou (1997) for different parameter values drawn from the true preference distribution. For these comparisons, we only use the designs obtained by means of the

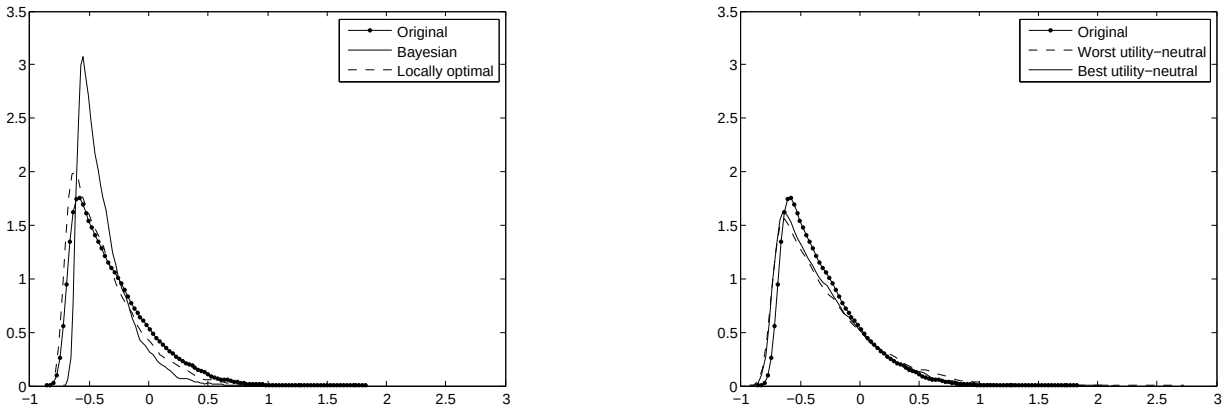


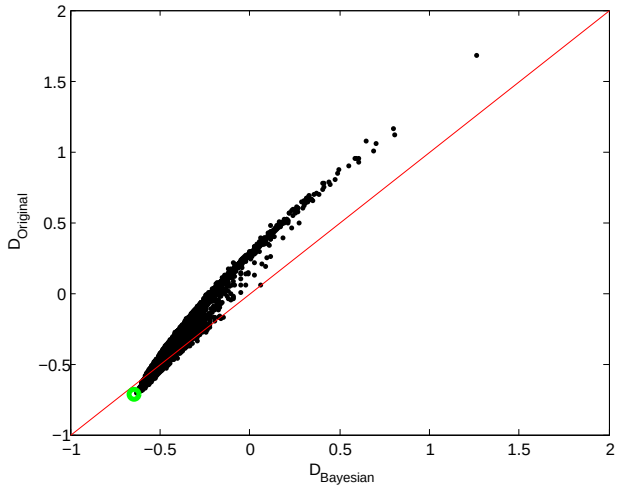
Figure 2.6: Distribution of D-optimality criterion values for the original, Bayesian, utility-neutral and locally optimal designs

coordinate-exchange algorithm. We again randomly draw $10,000$ parameter vectors from the $\mathcal{N}(\boldsymbol{\beta}_0, \Sigma_0)$ prior distribution and, for each draw, calculate D-optimality criterion values for all five designs. We visualize the resulting D-optimality criterion values in two ways. First, we plot the D-value distribution for each of the five designs. Second, we produce scatter plots of the D-optimality criterion values for the original design versus those for the Bayesian, locally optimal and utility-neutral designs.

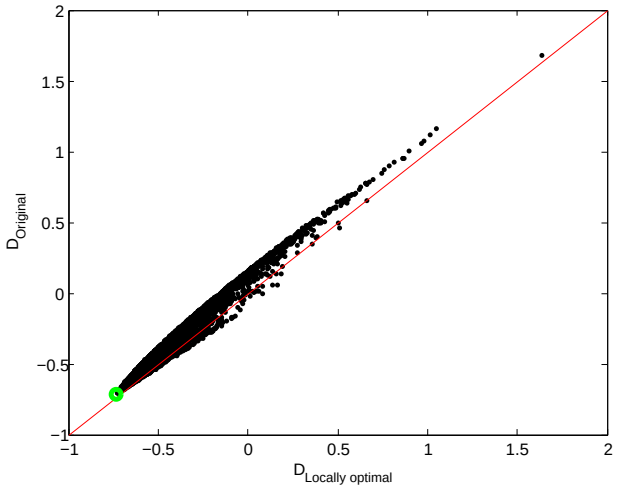
Figure 2.6 shows the estimated densities of the D-optimality criterion values for the original, Bayesian and locally optimal designs (left) and for the original and two utility-neutral designs (right). The density corresponding to the original experimental design has a substantially fatter and longer right tail than the densities for the Bayesian and locally optimal designs. This indicates that the original design can result in much higher (worse) D-values than the Bayesian or locally optimal design. The densities corresponding to both utility-neutral designs have right tails comparable to that of the original design. The densities also show that, for some parameter values, the original design yields smaller D-values than the Bayesian design. In turn, there are parameter values for which the locally optimal, Bayesian and both utility-neutral designs perform better than the original design.

Figure 2.7 shows scatter plots for (3,000 out of 10,000) D-values computed for the five designs. In each plot, we compare the original design to one alternative. Each plot shows the D-value for the original design on the vertical axis and the value for the alternative design on the horizontal axis. Points above the 45-degree line correspond to parameter vectors for which the alternative design produces a smaller D-value than the original design. Hence, larger fractions of points appearing above the 45-degree line signal a better relative perfor-

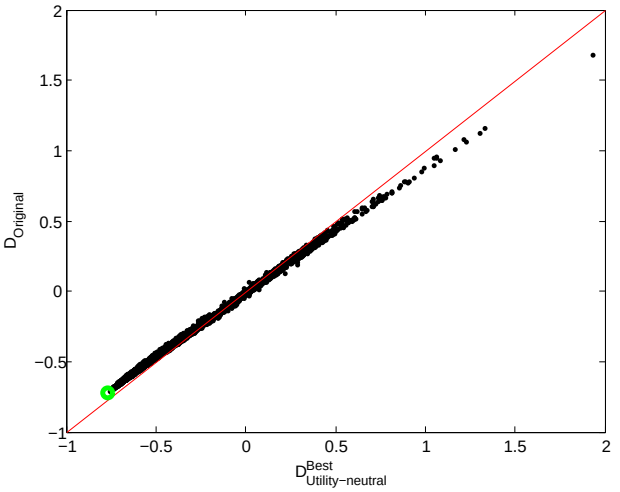
mance of the alternative design. The large green circle denotes the scenario corresponding to the zero parameter vector. The scatter plots show that the Bayesian design is substantially more robust than the original design. For the locally optimal design, this is the case as well, but the points are all closer to the 45-degree line. This is even more so for the utility-neutral designs. Note that whenever a point is below the 45-degree line for the Bayesian and locally optimal designs, it is never far below. Conversely, some of the points are well above the line. The Bayesian design outperforms the original design in 64.8% of the cases, the locally optimal design in 89.6% of the cases, the best utility-neutral design in 70.5% of the cases and the worst utility-neutral design in 56.2% of the cases. The locally optimal and best utility-neutral designs, thus, outperform the original design more often than the Bayesian design. In most cases, however, the difference in D-value in favor of the locally optimal and best utility-neutral designs is minor, whereas the difference in D-value in favor of the Bayesian design generally is substantially larger.



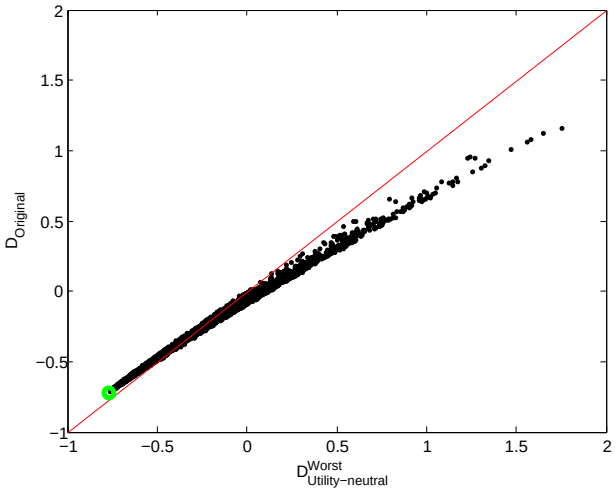
(a) Bayesian optimal design



(b) Locally optimal design



(c) Best utility-neutral design



(d) Worst utility-neutral design

Figure 2.7: Scatter plots of D-values for the original design used by Courcoux and Séménou (1997) and the Bayesian, locally optimal or utility-neutral designs

2.7 Conclusion

In this chapter, we studied the problem of constructing D-optimal experimental designs for the multinomial logit model when the alternatives are mixtures of ingredients. We developed two algorithms to find optimal designs for any number of ingredients and any type of Scheffé model: a mixture coordinate-exchange algorithm and a particle swarm optimization (PSO) algorithm. Because of the nonlinearity of the multinomial logit model, we adopt a Bayesian approach involving a prior distribution for the model parameters.

The PSO algorithm turns out to be faster than the mixture coordinate-exchange algorithm. However, in all settings we studied, designs produced by the PSO algorithm performed worse than designs constructed by using the mixture coordinate-exchange algorithm. We suspect that the reason why the PSO algorithm performs worse than the mixture coordinate-exchange algorithm is that particles quickly converge to local optima from which they cannot escape, as discussed in Clerc and Kennedy (2002).

We showed that the Bayesian and locally D-optimal designs differ both in look and in performance from the utility-neutral designs. Bayesian designs are robust in the sense that they perform well in a broad range of scenarios and that they never seem to perform very poorly. The Bayesian and locally D-optimal designs we construct are close to being utility balanced, which is considered a desirable feature in the literature on choice experiments.

Like all algorithms for constructing Bayesian optimal choice designs, the mixture coordinate-exchange algorithm which performed best for the design problem studied here is a heuristic. Therefore, it might not yield the globally optimal design. The extensive use of coordinate-exchange algorithms in the optimal experimental design literature, however, indicates that the resulting designs are highly efficient and practically useful. Nevertheless, an interesting avenue for future studies is to look for alternative optimization methods for mixtures. Furthermore, since the PSO requires less computing time to reach a reasonably good design, the PSO optimal design could be used as a starting design for the mixture coordinate-exchange algorithm. Additional computations show that for some of the experiments considered in this chapter, the combination of the PSO and the mixture coordinate-exchange algorithms led to a slightly better design in a relatively short computation time. However, it is not clear whether the mixture coordinate-exchange algorithm will always be capable of escaping from the locally optimal design obtained by the PSO. We believe that additional research is required to obtain the ideal algorithm. Another issue worth mentioning is that, while we focused on the D-optimality criterion, our algorithm can be adapted for generating designs that are optimal with respect to other criteria as well, such as the A-, G- and V-optimality criteria.

Finally, it is our hope that our work will have a beneficial impact on the use of choice experiments for mixtures. The examples in the textbook by Cornell (2002) suggest that ranking data are very commonly used, but it is known that ranking large numbers of products is tough for respondents. It is therefore better to use choice sets of size two (i.e., paired

comparisons) or three and ask the respondents to pick the mixture of their choice in each set (David, 1963). Our methods can guide the design of the experiments in these settings.

2.A Experimental designs

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.64	0.36	0.00	0.00	1.00
0.00	0.00	1.00	0.43	0.24	0.33
0.00	1.00	0.00	0.26	0.40	0.34
0.62	0.38	0.00	0.63	0.00	0.37
0.39	0.00	0.61	0.38	0.62	0.00
1.00	0.00	0.00	1.00	0.00	0.00
0.00	0.37	0.63			
0.00	1.00	0.00			

Table A.1: Optimal utility-neutral design with 14 alternatives in 7 choice sets of size 2, produced by PSO algorithm, for the special-cubic model for 3 ingredients. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.00	0.62	0.38
0.28	0.48	0.25	0.00	0.00	1.00
0.65	0.35	0.00	0.38	0.00	0.62
0.00	1.00	0.00	1.00	0.00	0.00
0.00	0.40	0.60	0.25	0.26	0.49
0.45	0.33	0.22	0.33	0.67	0.00
0.61	0.00	0.39			
0.00	0.00	1.00			

Table A.2: Optimal utility-neutral design with 14 alternatives in 7 choice sets of size 2, produced by mixture coordinate-exchange algorithm, for the special-cubic model for 3 ingredients. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.40	0.60	0.40	0.00	0.60
0.44	0.34	0.22	1.00	0.00	0.00
0.32	0.68	0.00	0.60	0.00	0.40
0.25	0.25	0.50	0.00	0.00	1.00
0.00	0.60	0.40	1.00	0.00	0.00
0.00	0.00	1.00	0.27	0.45	0.27
0.63	0.37	0.00			
0.00	1.00	0.00			

Table A.3: Optimal utility-neutral design with 14 alternatives in 7 choice sets of size 2, produced by JMP10, for the special-cubic model for 3 ingredients. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.00	0.63	0.37
0.37	0.63	0.00	0.39	0.22	0.38
1.00	0.00	0.00	0.60	0.40	0.00
0.35	0.00	0.65	0.00	0.60	0.40
0.00	0.24	0.76	0.00	0.00	1.00
0.00	1.00	0.00	0.36	0.32	0.33
0.26	0.44	0.30			
0.60	0.00	0.40			

Table A.4: Locally optimal design produced by mixture coordinate-exchange algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_1$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.22	0.46	0.33	1.00	0.00	0.00
0.59	0.00	0.41	0.38	0.00	0.62
0.25	0.29	0.47	0.00	0.71	0.29
0.00	1.00	0.00	0.00	0.00	1.00
0.00	0.35	0.65	0.00	1.00	0.00
0.49	0.25	0.26	0.63	0.37	0.00
1.00	0.00	0.00			
0.38	0.62	0.00			

Table A.5: Locally optimal design produced by mixture coordinate-exchange algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_2$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.55	0.00	0.45	0.00	1.00	0.00
0.00	0.57	0.43	0.00	0.00	1.00
0.00	0.45	0.55	0.32	0.26	0.41
0.00	1.00	0.00	0.84	0.16	0.00
0.00	0.00	1.00	0.43	0.57	0.00
1.00	0.00	0.00	1.00	0.00	0.00
0.43	0.00	0.57			
0.42	0.39	0.19			

Table A.6: Locally optimal design produced by PSO algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_1$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.66	0.34	0.26	0.42	0.32
0.38	0.26	0.36	0.00	0.00	1.00
0.62	0.00	0.38	1.00	0.00	0.00
0.00	0.00	1.00	0.38	0.00	0.62
0.00	0.50	0.50	0.00	1.00	0.00
0.62	0.38	0.00	0.00	0.29	0.71
0.38	0.62	0.00			
1.00	0.00	0.00			

Table A.7: Locally optimal design produced by PSO algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_2$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.23	0.44	0.33	1.00	0.00	0.00
0.58	0.00	0.42	0.41	0.00	0.59
0.23	0.30	0.48	0.00	0.68	0.32
0.00	1.00	0.00	0.00	0.00	1.00
0.00	0.33	0.67	0.00	1.00	0.00
0.47	0.26	0.27	0.60	0.40	0.00
1.00	0.00	0.00			
0.41	0.59	0.00			

Table A.8: Bayesian optimal design produced by mixture coordinate-exchange algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the prior $\mathcal{N}(\boldsymbol{\beta}_2, \kappa I_p)$ with $\kappa = 0.5$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.36	0.24	0.40	0.40	0.35	0.24
0.64	0.36	0.00	0.00	0.76	0.24
0.00	0.00	1.00	0.40	0.00	0.60
0.00	0.53	0.47	0.35	0.38	0.27
0.00	0.00	1.00	0.00	1.00	0.00
0.46	0.00	0.54	0.00	0.46	0.54
1.00	0.00	0.00			
0.54	0.46	0.00			

Table A.9: Bayesian optimal design produced by mixture coordinate-exchange algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the prior $\mathcal{N}(\boldsymbol{\beta}_2, \kappa I_p)$ with $\kappa = 5$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.40	0.26	0.34	0.39	0.00	0.61
0.61	0.39	0.00	0.00	0.00	1.00
0.41	0.00	0.59	0.00	0.44	0.56
0.38	0.33	0.28	0.00	0.00	1.00
0.00	0.54	0.46	0.63	0.37	0.00
0.00	1.00	0.00	1.00	0.00	0.00
0.36	0.42	0.22			
0.00	0.79	0.21			

Table A.10: Bayesian optimal design produced by mixture coordinate-exchange algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the prior $\mathcal{N}(\boldsymbol{\beta}_2, \kappa I_p)$ with $\kappa = 10$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.42	0.31	0.27	0.49	0.00	0.51
0.49	0.51	0.00	0.41	0.25	0.34
0.13	0.18	0.69	0.27	0.73	0.00
0.00	0.00	1.00	0.00	1.00	0.00
0.74	0.12	0.13	0.00	0.67	0.33
1.00	0.00	0.00	0.00	1.00	0.00
0.25	0.33	0.42			
0.00	0.50	0.50			

Table A.11: Bayesian optimal design produced by mixture coordinate-exchange algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the prior $\mathcal{N}(\boldsymbol{\beta}_2, \kappa I_p)$ with $\kappa = 30$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.47	0.27	0.27	0.24	0.47	0.29
0.00	0.76	0.24	0.50	0.00	0.50
0.00	1.00	0.00	1.00	0.00	0.00
0.00	0.33	0.67	0.00	0.00	1.00
0.00	0.66	0.34	0.37	0.16	0.47
0.00	0.00	1.00	0.63	0.37	0.00
1.00	0.00	0.00			
0.41	0.59	0.00			

Table A.12: Bayesian optimal design produced by PSO algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the prior $\mathcal{N}(\boldsymbol{\beta}_2, \kappa I_p)$ with $\kappa = 0.5$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.00	1.00	0.00	0.86	0.14
0.00	0.52	0.48	0.50	0.50	0.00
0.41	0.24	0.35	0.00	1.00	0.00
0.37	0.63	0.00	0.00	0.46	0.54
0.50	0.00	0.50	0.53	0.00	0.47
0.33	0.33	0.33	1.00	0.00	0.00
0.00	0.00	1.00			
0.50	0.00	0.50			

Table A.13: Bayesian optimal design produced by PSO algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the prior $\mathcal{N}(\boldsymbol{\beta}_2, \kappa I_p)$ with $\kappa = 5$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.33	0.33	0.33	0.42	0.58	0.00
0.57	0.00	0.43	0.39	0.27	0.34
1.00	0.00	0.00	0.00	0.24	0.76
0.64	0.00	0.37	0.34	0.25	0.40
0.38	0.62	0.00	0.00	0.53	0.47
0.00	1.00	0.00	0.00	1.00	0.00
0.00	0.44	0.56			
0.00	0.00	1.00			

Table A.14: Bayesian optimal design produced by PSO algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the prior $\mathcal{N}(\boldsymbol{\beta}_2, \kappa I_p)$ with $\kappa = 10$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.29	0.71	1.00	0.00	0.00
0.00	0.00	1.00	0.78	0.00	0.22
0.22	0.42	0.36	0.00	0.65	0.35
0.00	0.50	0.50	0.00	1.00	0.00
0.37	0.41	0.23	0.29	0.71	0.00
0.50	0.50	0.00	0.00	1.00	0.00
0.72	0.28	0.00			
0.86	0.00	0.14			

Table A.15: Bayesian optimal design produced by PSO algorithm with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the prior $\mathcal{N}(\boldsymbol{\beta}_2, \kappa I_p)$ with $\kappa = 30$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.00	1.00	0.00	0.44	0.56
0.00	0.59	0.41	0.00	1.00	0.00
0.27	0.32	0.41	0.00	0.53	0.47
0.50	0.50	0.00	0.39	0.33	0.28
0.00	0.53	0.47	0.00	1.00	0.00
0.39	0.33	0.28	0.00	0.44	0.56
0.30	0.39	0.31	0.00	0.00	1.00
0.49	0.00	0.51	0.58	0.00	0.42
1.00	0.00	0.00	0.50	0.50	0.00
0.48	0.52	0.00	0.27	0.32	0.41
0.00	0.00	1.00	0.00	1.00	0.00
0.58	0.00	0.42	0.52	0.48	0.00
0.54	0.46	0.00	1.00	0.00	0.00
0.48	0.00	0.52	0.43	0.00	0.57
0.00	0.00	1.00	0.43	0.00	0.57
0.30	0.38	0.31	1.00	0.00	0.00

Table A.16: Bayesian optimal design obtained by mixture coordinate-exchange algorithm using an uninformative prior with 32 alternatives in 16 choice sets of size 2 assuming the special-cubic model for the example from Courcoux and Séménou (1997), expressed in terms of the pseudo components. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.00	1.00	0.00
0.53	0.47	0.00	0.00	0.48	0.52
0.32	0.39	0.29	0.51	0.49	0.00
0.52	0.00	0.48	0.29	0.31	0.40
0.00	1.00	0.00	0.48	0.52	0.00
0.31	0.36	0.33	0.00	0.00	1.00
0.47	0.00	0.53	0.00	0.48	0.52
1.00	0.00	0.00	0.00	1.00	0.00
0.00	0.48	0.52	0.00	0.00	1.00
0.39	0.33	0.28	0.31	0.34	0.35
0.57	0.00	0.43	0.47	0.00	0.53
0.00	0.00	1.00	1.00	0.00	0.00
0.47	0.00	0.53	0.00	0.00	1.00
1.00	0.00	0.00	0.00	0.55	0.45
0.00	1.00	0.00	0.00	1.00	0.00
1.00	0.00	0.00	0.46	0.54	0.00

Table A.17: Bayesian optimal design obtained by PSO algorithm using an uninformative prior with 32 alternatives in 16 choice sets of size 2 assuming the special-cubic model for the example from Courcoux and Séménou (1997), expressed in terms of the pseudo components. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.50	0.50	1.00	0.00	0.00
0.36	0.32	0.32	0.50	0.50	0.00
0.50	0.00	0.50	0.00	0.40	0.60
0.30	0.38	0.33	0.44	0.00	0.56
0.48	0.52	0.00	1.00	0.00	0.00
0.32	0.33	0.35	0.50	0.50	0.00
0.55	0.00	0.45	0.35	0.32	0.32
1.00	0.00	0.00	1.00	0.00	0.00
0.45	0.00	0.55	0.00	0.48	0.52
0.00	0.00	1.00	0.00	1.00	0.00
0.00	0.47	0.53	0.00	1.00	0.00
0.00	0.00	1.00	0.29	0.38	0.33
0.00	1.00	0.00	0.36	0.32	0.32
0.51	0.49	0.00	0.00	0.50	0.50
0.51	0.49	0.00	0.00	0.47	0.53
0.00	1.00	0.00	0.00	0.00	1.00

Table A.18: Bayesian optimal design produced by mixture coordinate-exchange algorithm with 32 alternatives in 16 choice sets of size 2 for the example from Courcoux and Séménou (1997) assuming the special-cubic model and using a prior of $\mathcal{N}(\boldsymbol{\beta}_0, \Sigma_0)$, expressed in terms of the pseudo components. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	1.00	0.00	0.00	0.00	1.00
0.00	0.00	1.00	1.00	0.00	0.00
0.45	0.55	0.00	0.47	0.53	0.00
0.34	0.34	0.32	0.00	1.00	0.00
0.00	1.00	0.00	0.00	1.00	0.00
0.53	0.00	0.47	1.00	0.00	0.00
0.44	0.00	0.56	0.00	0.56	0.44
0.00	0.41	0.59	0.00	1.00	0.00
1.00	0.00	0.00	1.00	0.00	0.00
0.52	0.48	0.00	0.58	0.00	0.42
1.00	0.00	0.00	0.00	0.44	0.56
0.39	0.32	0.29	0.00	0.00	1.00
0.50	0.00	0.50	0.00	0.00	1.00
0.32	0.34	0.34	0.42	0.00	0.58
0.00	1.00	0.00	0.35	0.32	0.33
0.47	0.53	0.00	0.00	0.54	0.46

Table A.19: Bayesian optimal design produced by PSO algorithm with 32 alternatives in 16 choice sets of size 2 for the example from Courcoux and Séménou (1997) assuming the special-cubic model and using a prior of $\mathcal{N}(\boldsymbol{\beta}_0, \Sigma_0)$, expressed in terms of the pseudo components. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.35	0.31	0.34	0.63	0.37	0.00
0.00	1.00	0.00	0.00	1.00	0.00
0.00	0.37	0.63	0.41	0.30	0.29
0.00	1.00	0.00	0.00	0.47	0.53
0.00	0.00	1.00	0.00	0.64	0.36
1.00	0.00	0.00	0.00	0.00	1.00
0.36	0.64	0.00	0.47	0.00	0.53
1.00	0.00	0.00	0.31	0.42	0.28
0.40	0.00	0.60	0.00	1.00	0.00
1.00	0.00	0.00	0.50	0.00	0.50
0.52	0.00	0.48	0.00	0.00	1.00
0.00	0.49	0.51	0.47	0.53	0.00
0.54	0.46	0.00	1.00	0.00	0.00
0.00	0.52	0.48	0.29	0.37	0.34
0.60	0.00	0.40	0.31	0.28	0.41
0.00	0.00	1.00	0.50	0.50	0.00

Table A.20: Optimal utility-neutral experimental design obtained by mixture coordinate-exchange algorithm with 32 alternatives in 16 choice sets of size 2 assuming the special-cubic model for the example from Courcoux and Séménou (1997), expressed in terms of the pseudo components. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.00	1.00	0.00	0.53	0.47
0.54	0.46	0.00	0.39	0.29	0.32
0.64	0.00	0.36	0.00	0.00	1.00
0.00	0.00	1.00	0.00	0.52	0.48
0.46	0.54	0.00	0.00	0.00	1.00
0.00	0.51	0.49	0.00	1.00	0.00
0.37	0.00	0.63	0.51	0.49	0.00
1.00	0.00	0.00	0.55	0.00	0.45
0.27	0.40	0.33	0.00	0.42	0.58
0.47	0.00	0.53	0.00	1.00	0.00
0.00	1.00	0.00	0.45	0.55	0.00
0.53	0.00	0.47	1.00	0.00	0.00
1.00	0.00	0.00	1.00	0.00	0.00
0.31	0.31	0.38	0.00	1.00	0.00
0.31	0.33	0.35	0.00	1.00	0.00
0.00	0.00	1.00	0.59	0.41	0.00

Table A.21: Optimal utility-neutral experimental design obtained by PSO algorithm with 32 alternatives in 16 choice sets of size 2 assuming the special-cubic model for the example from Courcoux and Séménou (1997), expressed in terms of the pseudo components. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.31	0.35	0.34	0.37	0.63	0.00
1.00	0.00	0.00	1.00	0.00	0.00
0.37	0.00	0.63	0.30	0.41	0.29
1.00	0.00	0.00	0.47	0.00	0.53
0.00	0.00	1.00	0.64	0.00	0.36
0.00	1.00	0.00	0.00	0.00	1.00
0.64	0.36	0.00	0.00	0.47	0.53
0.00	1.00	0.00	0.42	0.31	0.28
0.00	0.40	0.60	1.00	0.00	0.00
0.00	1.00	0.00	0.00	0.50	0.50
0.00	0.52	0.48	0.00	0.00	1.00
0.49	0.00	0.51	0.53	0.47	0.00
0.46	0.54	0.00	0.00	1.00	0.00
0.52	0.00	0.48	0.37	0.29	0.34
0.00	0.60	0.40	0.28	0.31	0.41
0.00	0.00	1.00	0.50	0.50	0.00

Table A.22: Best utility-neutral experimental design for the study of quantifying cocktail preferences. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.34	0.31	0.35	0.00	0.37	0.63
0.00	1.00	0.00	0.00	1.00	0.00
0.63	0.37	0.00	0.29	0.30	0.41
0.00	1.00	0.00	0.53	0.47	0.00
1.00	0.00	0.00	0.36	0.64	0.00
0.00	0.00	1.00	1.00	0.00	0.00
0.00	0.64	0.36	0.53	0.00	0.47
0.00	0.00	1.00	0.28	0.42	0.31
0.60	0.00	0.40	0.00	1.00	0.00
0.00	0.00	1.00	0.50	0.00	0.50
0.48	0.00	0.52	1.00	0.00	0.00
0.51	0.49	0.00	0.00	0.53	0.47
0.00	0.46	0.54	0.00	0.00	1.00
0.48	0.52	0.00	0.34	0.37	0.29
0.40	0.00	0.60	0.41	0.28	0.31
1.00	0.00	0.00	0.00	0.50	0.50

Table A.23: Worst utility-neutral experimental design for the study of quantifying cocktail preferences. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	1.00	0.00	1.00	0.00	0.00
0.00	0.41	0.59	0.00	1.00	0.00
1.00	0.00	0.00	1.00	0.00	0.00
0.33	0.37	0.30	0.40	0.00	0.60
0.61	0.00	0.39	0.29	0.42	0.29
0.00	0.00	1.00	0.48	0.00	0.52
0.00	1.00	0.00	0.00	0.59	0.41
0.60	0.40	0.00	0.00	0.00	1.00
0.29	0.31	0.40	0.61	0.00	0.39
0.50	0.50	0.00	0.00	0.00	1.00
0.42	0.31	0.27	0.38	0.31	0.31
0.00	0.49	0.51	0.00	1.00	0.00
0.46	0.00	0.54	1.00	0.00	0.00
0.00	0.55	0.45	0.42	0.58	0.00
0.29	0.31	0.40	1.00	0.00	0.00
0.50	0.50	0.00	0.00	0.50	0.50

Table A.24: Locally optimal design obtained by mixture coordinate-exchange algorithm for the parameter vector $\boldsymbol{\beta}_0$ with 32 alternatives in 16 choice sets of size 2 assuming the special-cubic model for the example from Courcoux and Séménou (1997), expressed in terms of the pseudo components. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.30	0.32	0.38	0.00	0.50	0.50
0.50	0.50	0.00	0.51	0.00	0.49
0.00	1.00	0.00	0.29	0.40	0.32
0.00	0.00	1.00	0.55	0.00	0.45
0.00	0.00	1.00	0.00	1.00	0.00
0.56	0.00	0.44	0.35	0.32	0.33
0.00	0.00	1.00	1.00	0.00	0.00
1.00	0.00	0.00	0.43	0.57	0.00
0.00	0.53	0.47	0.00	1.00	0.00
1.00	0.00	0.00	1.00	0.00	0.00
0.00	0.44	0.56	0.00	0.00	1.00
0.00	1.00	0.00	0.00	0.56	0.44
0.00	1.00	0.00	1.00	0.00	0.00
0.61	0.39	0.00	0.50	0.50	0.00
0.46	0.00	0.54	0.00	0.50	0.50
1.00	0.00	0.00	0.39	0.30	0.30

Table A.25: Locally optimal design obtained by PSO algorithm for the parameter vector $\boldsymbol{\beta}_0$ with 32 alternatives in 16 choice sets of size 2 assuming the special-cubic model for the example from Courcoux and Séménou (1997), expressed in terms of the pseudo components. Choice sets are separated by dashed lines

2.B Permutations of the utility-neutral design

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.00	0.62	0.38
0.28	0.48	0.25	0.00	0.00	1.00
0.65	0.35	0.00	0.38	0.00	0.62
0.00	1.00	0.00	1.00	0.00	0.00
0.00	0.40	0.60	0.25	0.26	0.49
0.45	0.33	0.22	0.33	0.67	0.00
0.61	0.00	0.39			
0.00	0.00	1.00			

Table B.1: Permutation N^o 1 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	1.00	0.00	0.38	0.00	0.62
0.25	0.28	0.48	1.00	0.00	0.00
0.00	0.65	0.35	0.62	0.38	0.00
0.00	0.00	1.00	0.00	1.00	0.00
0.60	0.00	0.40	0.49	0.25	0.26
0.22	0.45	0.33	0.00	0.33	0.67
0.39	0.61	0.00			
1.00	0.00	0.00			

Table B.2: Permutation N^o 2 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.00	1.00	0.62	0.38	0.00
0.48	0.25	0.28	0.00	1.00	0.00
0.35	0.00	0.65	0.00	0.62	0.38
1.00	0.00	0.00	0.00	0.00	1.00
0.40	0.60	0.00	0.26	0.49	0.25
0.33	0.22	0.45	0.67	0.00	0.33
0.00	0.39	0.61			
0.00	1.00	0.00			

Table B.3: Permutation N^o 3 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.00	1.00	0.38	0.62	0.00
0.25	0.48	0.28	1.00	0.00	0.00
0.00	0.35	0.65	0.62	0.00	0.38
0.00	1.00	0.00	0.00	0.00	1.00
0.60	0.40	0.00	0.49	0.26	0.25
0.22	0.33	0.45	0.00	0.67	0.33
0.39	0.00	0.61			
1.00	0.00	0.00			

Table B.4: Permutation N^{o} 4 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	1.00	0.00	0.62	0.00	0.38
0.48	0.28	0.25	0.00	0.00	1.00
0.35	0.65	0.00	0.00	0.38	0.62
1.00	0.00	0.00	0.00	1.00	0.00
0.40	0.00	0.60	0.26	0.25	0.49
0.33	0.45	0.22	0.67	0.33	0.00
0.00	0.61	0.39			
0.00	0.00	1.00			

Table B.5: Permutation N^{o} 5 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.00	0.38	0.62
0.28	0.25	0.48	0.00	1.00	0.00
0.65	0.00	0.35	0.38	0.62	0.00
0.00	0.00	1.00	1.00	0.00	0.00
0.00	0.60	0.40	0.25	0.49	0.26
0.45	0.22	0.33	0.33	0.00	0.67
0.61	0.39	0.00			
0.00	1.00	0.00			

Table B.6: Permutation N^{o} 6 of the utility neutral design. Choice sets are separated by dashed lines

2.C Pseudo code for the PSO algorithm

Notation:

$D^t \coloneqq$ minimum D-value found at iteration t,

$D_i^t \coloneqq$ D-value of particle i at iteration t,

$\mathrm{iterMax} \coloneqq$ maximum number of iterations t,

$\text{maxProportion}_j \coloneqq$ maximum proportion of alternative j,

$n \coloneqq$ the number of alternatives in the experimental design,

$q \coloneqq$ the number of ingredient proportions,

$p \coloneqq$ the number of parameters in the model,

$S \coloneqq$ the number of choice sets,

$N \coloneqq$ the number of particles,

$K \coloneqq$ the number of elements in a particle, $K = n \times q$,

$\mathbf{v}_i^t \coloneqq$ velocity of particle i at iteration t, with elements v_{ki}^t,

$w_t \coloneqq$ inertia weight at iteration t,

$c_1 \coloneqq$ cognitive learning factor,

$c_2 \coloneqq$ social learning factor,

$\chi \coloneqq$ constriction factor,

$\boldsymbol{\xi}_i^t \coloneqq$ particle i at iteration t over a simplex,

$\tilde{\boldsymbol{\xi}}_i^t \coloneqq$ particle i at iteration t over a unit hypercube, with elements $\tilde{\xi}_{ki}^t$,

$\tilde{\boldsymbol{\xi}}_i^{pbest} \coloneqq$ design that has a minimum D-value for a particle i at a particular iteration, i.e., personal best, over a unit hypercube, with elements $\tilde{\xi}_{ki}^{pbest}$,

$\tilde{\boldsymbol{\xi}}^{gbest} \coloneqq$ design that has a minimum D-value found so far (D^t) by any of the particles, i.e., global best, over a unit hypercube, with elements $\tilde{\xi}_k^{gbest}$.

Then, the PSO algorithm for constructing the D-optimal experimental design can be described as follows:

for $i \leftarrow 1,\ N$ **do**

$\boldsymbol{\xi}_i^0 \leftarrow$ OBTAINSTARTINGDESIGN(n, q, p, S)

$\tilde{\boldsymbol{\xi}}_i^0 \leftarrow$ TRANSFORMTOHYPERCUBE($\boldsymbol{\xi}_i^0$)

$D_i \leftarrow \mathrm{D}(\boldsymbol{\xi}_i^0)$

$\mathbf{v}_i \leftarrow \mathbf{0}$

end for

$\tilde{\boldsymbol{\xi}}_i^{pbest} \leftarrow \tilde{\boldsymbol{\xi}}_i^0$

$j \leftarrow \arg\min_i D_i$

$\tilde{\boldsymbol{\xi}}^{gbest} \leftarrow \tilde{\boldsymbol{\xi}}_j^0$

$D^0 \leftarrow D_j$

$D^1 \leftarrow 10 + D^0$

$t \leftarrow 0$

while $(\text{abs}(D^{t+1} - D^t) > 0.0001) \wedge (t < \text{iterMax})$ **do**

$t \leftarrow t + 1$

$D^t \leftarrow D^{t-1}$

$\epsilon_k^1 \leftarrow \text{rand}(0, 1)\ \forall k = 1, \ldots, K$

$\epsilon_k^2 \leftarrow \text{rand}(0, 1)\ \forall k = 1, \ldots, K$

for $i \leftarrow 1,\ N$ **do**

for $k \leftarrow 1,\ K$ **do**

$v_{ki}^t \leftarrow w_t v_{ki}^{t-1} + c_1 \epsilon_k^1 (\tilde{\xi}_k^{gbest} - \tilde{\xi}_{ki}^{t-1}) + c_2 \epsilon_k^2 (\tilde{\xi}_{ki}^{pbest} - \tilde{\xi}_{ki}^{t-1})$

if $v_{ki}^t > 1$ **then**

$v_{ki}^t \leftarrow 1$

end if

if $v_{ki}^t < -1$ **then**

$v_{ki}^t \leftarrow -1$

end if

$\tilde{\xi}_{ki}^t \leftarrow \tilde{\xi}_{ki}^{t-1} + \chi v_{ki}^t$

end for

$\boldsymbol{\xi}_i^t \leftarrow \text{P}(\tilde{\boldsymbol{\xi}}_{\boldsymbol{i}}^{\boldsymbol{t}})$

$D_i^t \leftarrow \text{D}(\boldsymbol{\xi}_i^t)$

if $\text{D}_i^t < \text{D}_i^{t-1}$ **then**

$\tilde{\boldsymbol{\xi}}_i^{pbest} \leftarrow \tilde{\boldsymbol{\xi}}_i^t$

end if

end for

$j \leftarrow \arg\min_i D_i^t\ \forall i = 1, \ldots, N$

$\tilde{\boldsymbol{\xi}}^{gbest} \leftarrow \tilde{\boldsymbol{\xi}}_j$

$D^{t+1} \leftarrow \text{D}_j$

end while

return $\boldsymbol{\xi} = \text{P}(\tilde{\boldsymbol{\xi}}^{gbest})$

procedure OBTAINSTARTINGDESIGN(n, q, p, S)

1. Choose p distinct design points that maximize the information matrix for a minimum support design with p observations (see Section 2.4.1 for a discussion)

2. Replicate them as evenly as possible to obtain n observations for the design matrix X

3. Spread the replicated design points as evenly as possible over the choice sets. Avoid replicating points within a choice set

4. Assign the remaining non-replicated points to choice sets.

end procedure

procedure TRANSFORMTOHYPERCUBE($\boldsymbol{\xi}$)

$\tilde{\boldsymbol{\xi}} \leftarrow \boldsymbol{\xi}$

for $j \leftarrow 1, n$ **do**

$u_j \leftarrow$ rand(0, $\frac{1}{\text{maxProportion}_j}$)

Multiply elements in $\tilde{\boldsymbol{\xi}}$ corresponding to alternative j by u_j

end for

return $\text{vec}(\tilde{\boldsymbol{\xi}})$

end procedure

procedure P($\tilde{\boldsymbol{\xi}}$)

$\boldsymbol{\xi} \leftarrow \tilde{\boldsymbol{\xi}}$

for $j \leftarrow 1, n$ **do**

Divide elements in $\boldsymbol{\xi}$ corresponding to alternative j by their sum

end for

Reshape $\boldsymbol{\xi}$ to a matrix of size $n \times q$, which contains n alternatives in rows

return $\boldsymbol{\xi}$

end procedure

procedure D($\cdot$)

Compute the D-value (see Equation (2.6))

end procedure

2.D Pseudo code for the mixture coordinate-exchange algorithm

Notation:

$X \coloneqq$ the design matrix,
$n \coloneqq$ the number of alternatives in the experimental design,
$q \coloneqq$ the number of ingredient proportions,
$p \coloneqq$ the number of the parameters in the model,
$S \coloneqq$ the number of choice sets,
$\mathbf{S} \coloneqq$ the set of starting designs,
$\mathbf{D}^S \coloneqq$ the set of D-values of starting designs in $\mathbf{S}$,
$\mathcal{S} \coloneqq$ a particular starting design with the D-value D^S,
$\mathbf{O} \coloneqq$ the set of the constructed optimal designs for each of the starting design,
$\mathbf{D}^{Opt} \coloneqq$ the set of D-values of optimized designs in $\mathbf{O}$,
$\mathcal{O} \coloneqq$ a particular optimal experimental design with the D-value D^{Opt}.

Then, the mixture coordinate-exchange algorithm for constructing the D-optimal experimental design can be described as follows:

parfor $i \leftarrow 1, \text{starts}$ **do**
  $\mathbf{S}(i) \leftarrow$ OBTAINSTARTINGDESIGN(n, q, p, S)
  $[\mathbf{O}(i), \mathbf{D}^{Opt}(i)] \leftarrow$ OPTIMIZESTARTINGDESIGN($\mathbf{S}(i)$)
end parfor
$i \leftarrow \arg\min_k \mathbf{D}^{Opt}(k)$
return $\mathbf{O}(i)$

procedure OBTAINSTARTINGDESIGN(n, q, p, S)

1. Choose p distinct design points that maximize the information matrix for a minimum support design with p observations (see Section 2.4.1 for a discussion)

2. Replicate them as evenly as possible to obtain n observations for the design matrix X

3. Spread the replicated design points as evenly as possible over the choice sets. Avoid replicating points within a choice set

4. Assign the remaining non-replicated points to choice sets.

end procedure

procedure OPTIMIZESTARTINGDESIGN($\mathcal{S}$)

$D^S \leftarrow \mathrm{D}(\mathcal{S})$

$D^{\mathrm{Init}} \leftarrow 10 + D^S$

$D^{\mathrm{Opt}} \leftarrow D^{\mathcal{S}}$

$X \leftarrow \mathcal{S}$

while $\mathrm{abs}(D^{\mathrm{Init}} - D^{\mathrm{Opt}}) > 0.0001$ **do**

$D^{\mathrm{Init}} \leftarrow D^{\mathrm{Opt}}$

for $i \leftarrow 1,\ n$ **do**

for $j \leftarrow 1,\ q$ **do**

$X^* \leftarrow$ BRENT(X, i, j)

$D^{\mathrm{New}} \leftarrow \mathrm{D}(X^*)$

if $D^{\mathrm{New}} < D^{\mathrm{Opt}}$ **then**

$D^{\mathrm{Opt}} \leftarrow D^{\mathrm{New}}$

$X \leftarrow X^*$

end if

end for

end for

end while

return X, D^{Opt}

end procedure

procedure D($\cdot$)

Compute the D-value (see Equation (2.6))

end procedure

procedure BRENT(X, i, j)

Optimize a coordinate (i, j) of a design matrix X by using the method of Brent (1973) (see Section 2.4.2)

return optimal design X^*

end procedure

Chapter 3

Flexible Mixture-Amount Models for Business and Industry Using Gaussian Processes

3.1 Introduction

Many products and services can be described as mixtures of ingredients. Examples are mixtures of different fruits composing a fruit salad (e.g. 50% of apples, 30% of wild berries and 20% of grapes) or the mixture of different transportation modes used by an individual on a particular trip (e.g. 70% of travel time by metro and 30% by bike). In marketing, advertisers have to decide on the advertising media mix (e.g. 30% of the expenditures on TV advertising, 10% on radio and 60% on the Internet). As another example, hormone mixture treatments are of interest in biological research. In general, the response to such a product, service, media mix or treatment depends on the proportions of the individual ingredients. To explain such responses, specialized models are necessary to account for the fact that the proportions sum to one (Cornell, 2002).

In many cases, some other quantitative variable describing each mixture may also be relevant, both to the effect of individual ingredients on the response and to the response itself. In the marketing example, advertisers decide on the advertising media mix as well as on the *total budget* of the entire campaign. The total advertising budget will of course affect the impact of the campaign. Additionally, it is likely that the total budget also affects the

impact of a particular advertising medium. In a transportation setting, the attractiveness of a trip depends on the mix of transportation modes but also on the *total travel time*. However, the total travel time can affect the sensitivity of the attractiveness to particular transportation modes as well. Finally, the choice of a salad is affected by both its ingredients and the price. At the same time, the price may have an impact on how important different ingredients composing the salad are.

In general, a quantitative variable often impacts not only the response but also the effect of each ingredient in a mixture. Although this quantitative variable does not always correspond to a true amount, for simplicity, we will refer to this variable as the amount variable. Models that simultaneously link mixture proportions and amount variables to response variables are called mixture-amount models (Piepel and Cornell, 1985; Cornell, 2002).

If the total amount of a mixture affects the impact of mixture proportions, the parameters corresponding to the mixture ingredients in a model need to vary with the amount. For this reason, mixture-amount models typically express the mixture parameters as a parametric function of the amount. The effect of the amount on the response is then captured through its effect on the mixture parameters (Piepel and Cornell, 1985). However, such models require the specification of a functional form relating the mixture parameters to the amount variable a priori. Correctly specifying such a function may not be straightforward. Some flexible functional forms are available, see Piepel and Cornell (1985). However, the number of parameters in these specifications is usually very large. This prevents the use of the resulting models in practice, as these models are usually fitted to experimental data, the sample size of which tends to be small.

In this chapter, we introduce an alternative approach which is parsimonious in the number of parameters as well as flexible. Our approach is based on so-called Gaussian processes (Rasmussen and Williams, 2005) and avoids the necessity to specify the shape of the functional form of the relationship between the amount variable and the mixture parameters. We only use a smoothness assumption, meaning that, for similar values of the amount, we expect the mixture parameters to be similar as well. The degree of smoothness is captured by a parameter that can also be estimated if sufficient data are available. Another way to interpret our model is that we treat mixture parameters as functions of the amount and that we specify a prior distribution directly over these functions.

In technical terms, we specify a separate parameter vector for every unique observed amount value. One such parameter vector describes the impact of the mixture components on the response at a specific amount. These parameter vectors are, however, not independent across amounts. As explained above, the model incorporates the idea that, for amount values that are close to each other, the model parameters are expected to be rather similar. The Gaussian process formalizes this by specifying the correlation between all parameter vectors. The correlation structure itself is governed by the so-called Gaussian kernel, which is described by a single parameter. This parameter specifies the dependence of the correlations on the amount differences and, therefore, controls for the smoothness of the mixture parameters as a function of the amount. If one sets this parameter to zero or to a large positive value, one can obtain existing models as special cases. When the parameter equals zero, the correlations approach zero and one obtains different and independent mixture parameters for each unique amount value. Such a model has been considered by Piepel and Cornell (1985). When the parameter approaches infinity, the correlations tend to one and one obtains a single vector for the mixture parameters such that the amount variable does not play a role. In this case, we are left with a standard mixture model, as, for example, used in Sahrmann et al. (1987).

Finally, apart from the correlations across amounts, the mixture parameters of the model at a given amount value might also be correlated. For instance, the impacts of radio advertising and TV advertising may move up and down together as one considers different advertising amounts. In our model, we also allow for this type of correlation. As a result, the overall variance-covariance structure of the mixture parameters depends on a parameter that controls the correlation across amounts and a parameter that controls the correlation across individual parameters at a given amount. If the latter correlation approaches one, we obtain another special case of our model in which the amount has a separate, additive impact on the response variable.

We demonstrate that our approach naturally leads to a model specification in which the mixture parameters follow a matrix normal distribution with the variance-covariance matrix consisting of two parts. The parameters of the resulting model can be estimated using Bayesian techniques. In this chapter, we also provide the details of the required sampling procedures.

To illustrate our approach, we present two examples. The first example concerns the reaction of mice to different mixtures of hormones administered at different amount levels. The second illustration considers the recognition of advertising campaigns for skin and hair care products. The mixture here is a particular media mix used for a campaign. The amount variable is the total advertising campaign exposure. We introduce both examples in more detail in the subsequent sections.

The remainder of this chapter is organized as follows. In the next section, we review the literature on mixture-amount models and Gaussian processes. Section 3.3 introduces our new approach to model mixture-amount data. Section 3.4 presents our Bayesian estimation procedure. In Section 3.5, we illustrate the new modeling approach. We end the chapter with a discussion in Section 3.6.

3.2 Literature

In this section, we first review the existing literature on mixture-amount models. Next, we discuss Gaussian processes which we use to develop our new models for mixture-amount data.

3.2.1 Mixture-amount models

When a response variable is modeled as a function of proportions of ingredients in a mixture, the *mixture constraint*, defined by $\sum_{i=1}^{q} x_i = 1$, has a significant impact on the models that can be fitted. Here, x_i is the proportion of ingredient i and q is the number of ingredients in the mixture. The first consequence is that a linear regression model for mixture data cannot contain an intercept. Furthermore, cross-products $x_i x_j$ and squares x_i^2 cannot be simultaneously included as regressors in the model, since this leads to perfect collinearity. To deal with these issues, Scheffé (1958, 1963) proposed a family of models that are suitable for modeling mixture data. The first-order (linear) and second-order Scheffé models, respectively, for a continuous dependent variable y are defined as

$$y = \sum_{i=1}^{q} \beta_i x_i + \varepsilon \tag{3.1}$$

and

$$y = \sum_{i=1}^{q} \beta_i x_i + \sum \sum_{i<j}^{q} \beta_{ij} x_i x_j + \varepsilon, \tag{3.2}$$

where ε indicates the error term.

The models in Equations (3.1)-(3.2) can be used if the total amount is fixed or does not affect the response. However, they are not suitable if the amount of a mixture affects the response. Piepel and Cornell (1985) introduced mixture-amount models to deal with situations in which the response depends on the total amount of a mixture as well as on the ingredient proportions. They recognized the similarity of a mixture-amount experiment to a mixture experiment with one process variable (the amount in this case) and adapted models developed by Scheffé (1963) for mixture experiments with process variables.

Following Piepel and Cornell (1985), assume that we have acquired mixture data at r different values of the amount variable A, denoted by $A_1, A_2, \ldots, A_r$ ($r \geq 2$), and that the relation between the response and the ingredient proportions is modeled by a Scheffé model with p mixture parameters, $\beta_1, \beta_2, \ldots, \beta_p$. If the total amount of the mixture affects the impact of mixture proportions, the parameters corresponding to the mixture ingredients in a model need to vary with A. Thus, each mixture parameter β_m, $m = 1, \ldots, p$, has to depend on the total amount. Using this reasoning, one can create a *mixture-amount model* from the assumed Scheffé model by allowing the mixture parameters β_m to be a function $\beta_m(A)$ of the amount A, for $m = 1, \ldots, p$. One possible parametric model for the dependence of the mixture parameters on the amount is the polynomial function,

$$\beta_m(A) = \beta_m^0 + \sum_{k=1}^{K} \beta_m^k A^k. \tag{3.3}$$

The parameter β_m^k represents the k^{th} order effect of the amount on β_m.

As an example, we present a model for mixture-amount data for $q = 2$ ingredients based on the second-order Scheffé model given in Equation (3.2) and using the expression in Equation (3.3) with $K = 2$ to write the mixture parameters as a function of the amount:

$$\begin{aligned} y &= \beta_1(A)x_1 + \beta_2(A)x_2 + \beta_3(A)x_1x_2 + \varepsilon \\ &= \beta_1^0 x_1 + \beta_2^0 x_2 + \beta_3^0 x_1 x_2 + \sum_{k=1}^{2} (\beta_1^k x_1 + \beta_2^k x_2 + \beta_3^k x_1 x_2) A^k + \varepsilon. \end{aligned} \tag{3.4}$$

This model contains first- and second-order effects of the mixture components and linear and quadratic effects of the amount variable. The terms in the mixture-amount model in Equation (3.4) have the following interpretation:

- if the amount variable A is centered around zero, $\beta_1^0 x_1 + \beta_2^0 x_2 + \beta_{12}^0 x_1 x_2$ represents the linear and nonlinear blending properties of the mixture components at the average value of the total amount;
- $(\beta_1^1 x_1 + \beta_2^1 x_2 + \beta_{12}^1 x_1 x_2)A$ represents the linear effect of the total amount on the linear and nonlinear blending properties of the mixture components;
- $(\beta_1^2 x_1 + \beta_2^2 x_2 + \beta_{12}^2 x_1 x_2)A^2$ represents the quadratic effect of the total amount on the linear and nonlinear blending properties of the mixture components.

In general, the parameters β_i^k and β_{ij}^k of the terms involving $x_i A^k$ and $x_i x_j A^k$ $(k = 1, 2)$ in Equation (3.4) are measures of the effect of changing the total amount of the mixture on the linear and nonlinear blending properties of the mixture ingredients. For general q and k, we have

$$y = \sum_{i=1}^{q} \beta_i^0 x_i + \sum\sum_{i<j}^{q} \beta_{ij}^0 x_i x_j + \sum_{k=1}^{K} \left[\sum_{i=1}^{q} \beta_i^k x_i + \sum\sum_{i<j}^{q} \beta_{ij}^k x_i x_j \right] A^k + \varepsilon. \tag{3.5}$$

To emphasize the fact that the mixture parameters are assumed to be some parametric functions of the amount, we call the models above parametric. When the amount of a mixture does not affect the blending properties of the mixture components but only causes a constant change in the magnitude of the response (that is, all β_i^k are equal and all $\beta_{ij}^k = 0$), Equation (3.5) reduces to

$$y = \sum_{i=1}^{q} \beta_i^0 x_i + \sum\sum_{i<j}^{q} \beta_{ij}^0 x_i x_j + \sum_{k=1}^{K} \beta_0^k A^k + \varepsilon, \tag{3.6}$$

where $\beta_0^k = \beta_1^k = \cdots = \beta_q^k$. In this case, the amount does not affect the impact of the proportions on the response, but it has a direct impact itself.

The models specified above are typically used for mixture-amount data. However, there are a number of issues with them. First, the number of parameters in the final model grows rapidly with q and K. Furthermore, K has to be specified a priori, which is not always easy to do. Third, using a large value for K may yield highly volatile functions $\beta_m(A)$. Finally, in

addition to polynomial functions, there is a wide variety of other specifications that one may want to consider. To avoid all these issues, in this chapter, we introduce a non-parametric specification for $\beta_m(A)$ based on Gaussian processes. This approach does not require an a priori selection of the shape of the functions $\beta_m(A)$.

Below, we first discuss Gaussian processes in general. In Section 3.3, we incorporate the Gaussian process in the mixture-amount model.

3.2.2 Gaussian processes

A Gaussian process (GP) defines a distribution over functions. Denote such a distribution by $P(f)$ for some function f, $f : \chi \to \mathbb{R}$. Then, $P(f)$ is a Gaussian process if for any finite subset of χ, the marginal distribution over that finite subset has a multivariate Gaussian distribution (Rasmussen and Williams, 2005; Bishop, 2006). We can therefore write $f(\mathbf{x}) \sim \mathcal{N}\big(m(\mathbf{x}), \mathbf{\Omega}(\mathbf{x}, \mathbf{x})\big)$, $\mathbf{x} \subset \chi$, for a mean function $m(\mathbf{x})$ and covariance function $\mathbf{\Omega}(\mathbf{x}, \mathbf{x})$. As a result, a Gaussian process is parameterized by its mean and covariance functions. Note that f can be infinite-dimensional and therefore Gaussian processes extend multivariate Gaussian distributions to infinite dimensionality.

After some mean is assumed for $f(\mathbf{x})$, the covariance function $\mathbf{\Omega}(\mathbf{x}, \mathbf{x})$ completely defines the behavior of $f(\mathbf{x})$ for different values of $\mathbf{x}$. The function $\mathbf{\Omega}(\mathbf{x}, \mathbf{x})$ parameterizes our beliefs about the smoothness of $f(\mathbf{x})$ with respect to $\mathbf{x}$. Different $\mathbf{\Omega}(\mathbf{x}, \mathbf{x})$ functions could represent many different kinds of nonlinearity and lead to different shapes of $f(\mathbf{x})$ (Rasmussen and Williams (2005), see also Wilson and Adams (2013); Salimans (2012); Duvenaud et al. (2013)). In general, any real-valued function $\mathbf{\Omega}(\mathbf{x}, \mathbf{x})$ is acceptable to describe a covariance function provided the resulting covariance matrix is positive semi-definite.

By estimating the parameters defining the mean and covariance functions of f, we in fact acquire knowledge concerning the distribution of f. Note that, in this process, we do not assume any parametric form for the function f itself. Prior beliefs about the structure of the function f can be incorporated by choosing a particular covariance function. As a result, Gaussian processes are very flexible and can be used to represent many different regression models that would have an infinite number of parameters if formulated in a conventional manner (Neal, 1999).

Prediction for Gaussian processes is easy if the mean and covariance functions are known. Suppose that we already know the function's values at $\mathbf{x}$ and wish to predict the function's value at a new observation $\mathbf{x}^*$, i.e., $f(\mathbf{x}^*)$. Recall that for any function f drawn from a Gaussian process prior with the mean and covariance functions given by $m(\cdot)$ and $\boldsymbol{\Omega}(\cdot,\cdot)$, respectively, the marginal distribution over any finite subset of χ is multivariate Gaussian. Therefore, the joint distribution of f at the observed data $\mathbf{x}$ and at the new data point $\mathbf{x}^*$ can be written as

$$\begin{pmatrix} f(\mathbf{x}) \\ f(\mathbf{x}^*) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m(\mathbf{x}) \\ m(\mathbf{x}^*) \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega}(\mathbf{x},\mathbf{x}) & \boldsymbol{\Omega}(\mathbf{x},\mathbf{x}^*) \\ \boldsymbol{\Omega}(\mathbf{x}^*,\mathbf{x}) & \boldsymbol{\Omega}(\mathbf{x}^*,\mathbf{x}^*) \end{pmatrix} \right),$$

where $m(\cdot)$ and $\boldsymbol{\Omega}(\cdot,\cdot)$ denote the mean and covariance functions evaluated at either the observed data $\mathbf{x}$ or at the new data $\mathbf{x}^*$. Conditioning the joint Gaussian prior distribution on the observations gives

$$\begin{aligned} f(\mathbf{x}^*)|\mathbf{x},\mathbf{x}^*,f(\mathbf{x}) \sim \mathcal{N}\big(& m(\mathbf{x}^*) + \boldsymbol{\Omega}(\mathbf{x}^*,\mathbf{x})\boldsymbol{\Omega}(\mathbf{x},\mathbf{x})^{-1}(f(\mathbf{x}) - m(\mathbf{x})), \\ & \boldsymbol{\Omega}(\mathbf{x}^*,\mathbf{x}^*) - \boldsymbol{\Omega}(\mathbf{x}^*,\mathbf{x})\boldsymbol{\Omega}(\mathbf{x},\mathbf{x})^{-1}\boldsymbol{\Omega}(\mathbf{x},\mathbf{x}^*)\big), \end{aligned} \tag{3.7}$$

which is the posterior predictive distribution of $f(\mathbf{x}^*)$ for any input $\mathbf{x}^*$. Function values $f(\mathbf{x}^*)$ can be sampled from the joint posterior distribution by evaluating the mean and covariance functions in Equation (3.7) (Rasmussen and Williams, 2005).

Gaussian processes are conceptually simple and flexible, and they often exhibit a good performance in various applications. Thus, it is not surprising that they are widespread in many different areas ranging from simple regressions and classifications (Neal, 1997, 1999; Williams, 1998; Gattiker et al., 2015) or multi-task learning (Melkumyan and Ramos, 2011; Bonilla et al., 2007; Boyle and Frean, 2005) to visualisation of high dimensional data (Lawrence, 2004), density estimation (Leonard, 1978; Riihimäki and Vehtari, 2014) or human motion modeling (Wang et al., 2008). However, Gaussian processes have hitherto not been used in the context of mixture-amount models.

3.3 Model

3.3.1 Derivation

As discussed above, a straightforward way to model mixture-amount data is to specify the dependence of mixture parameters on the amount explicitly, like in Equation (3.3). In this section, we present an elegant way to model $\beta_m,\ m = 1, \ldots, p$, as a function of the amount A using Gaussian processes (Rasmussen and Williams, 2005), where we do not explicitly assume any functional form.

We denote the set of observed amount values in the data as $\vec{A} = (A_1, A_2, \ldots, A_r)'$, $r \leq N$, where N is the total number of observations. The latent function linking the mixture parameters β_m to the amount is given by $\beta_m(A)$. Our approach specifies a (prior) distribution directly over these functions, where the correlation structure in $\mathbf{\Omega}(\cdot, \cdot)$ is specified using only one positive parameter (τ). This parameter determines how quickly the mixture parameters vary with respect to the amount.

Formally, we collect the parameters for all observed amount values in the parameter matrix $\mathbf{B}(\vec{A})$, which contains the p ingredient's and their interactions' effects at different amount values in its rows and different ingredient's and their interactions' effects at a given value of the amount in its columns, i.e.,

$$\mathbf{B}(\vec{A}) = \begin{pmatrix} \beta_1(A_1) & \beta_2(A_1) & \ldots & \beta_p(A_1) \\ \beta_1(A_2) & \beta_2(A_2) & \ldots & \beta_p(A_2) \\ \vdots & \vdots & \vdots & \vdots \\ \beta_1(A_r) & \beta_2(A_r) & \ldots & \beta_p(A_r) \end{pmatrix} = \begin{pmatrix} \beta_1(\vec{A})' \\ \beta_2(\vec{A})' \\ \vdots \\ \beta_p(\vec{A})' \end{pmatrix}', \tag{3.8}$$

with $\beta_m(\vec{A}) = \big(\beta_m(A_1), \beta_m(A_2), \ldots, \beta_m(A_r)\big)'$.

If we assume that the response variable is continuous and consider a linear Scheffé model, then $p = q$ and the response y_i of an observation i can be modeled as

$$y_i = \big(\mathbf{a}_i \mathbf{B}(\vec{A})\big) \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{qi} \end{pmatrix} + \varepsilon_i, \tag{3.9}$$

where $\mathbf{a}_i$ is a $1 \times r$ row vector indicating which of the amount values corresponds to observation i. The j^{th} element of $\mathbf{a}_i$ is one if the j^{th} amount is used for observation i and zero otherwise. The row vector $\mathbf{a}_i$ selects the appropriate parameters from $\mathbf{B}(\vec{A})$. Using some linear algebra, we can rewrite the model for y_i as

$$y_i = (\mathbf{x}_i' \otimes \mathbf{a}_i)\,\mathrm{vec}\big(\mathbf{B}(\vec{A})\big) + \varepsilon_i, \tag{3.10}$$

where $\otimes$ is the Kronecker product, $\mathrm{vec}(\cdot)$ denotes the vectorization operator and $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{qi})'$. Stacking all response values gives

$$\mathbf{y} = \mathbf{X}\mathrm{vec}\big(\mathbf{B}(\vec{A})\big) + \boldsymbol{\varepsilon}, \tag{3.11}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_N)'$, $\mathbf{X} = (X_1', X_2', \ldots, X_N')'$ with $X_i = \mathbf{x}_i' \otimes \mathbf{a}_i$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$.

The model in Equation (3.11) resembles a standard linear regression model. The only difference is that we treat the parameter vector $\mathrm{vec}\big(\mathbf{B}(\vec{A})\big)$ as a function of the (observed) amount values. To complete the model, we assume that the prior on the parameters $\beta_m(A)$ is a Gaussian process with mean b_m and covariance function $\boldsymbol{\Omega}(\cdot,\cdot)$. We also allow the Gaussian processes to be correlated across m, that is, we allow for correlation between the different mixture ingredient parameters. At a given amount level, the variance-covariance matrix of the p mixture parameters is given by $\sigma^2\boldsymbol{\Phi}$.

As a result, the parameter matrix at the observed amounts, $\mathbf{B}(\vec{A})$, follows a matrix-normal distribution, that is, $\mathbf{B}(\vec{A})|\sigma^2 \sim \mathcal{MN}\big(\bar{\mathbf{B}}, \boldsymbol{\Omega}, \sigma^2\boldsymbol{\Phi}\big)$, where

$$\bar{\mathbf{B}} = (\mathbf{1}_{r\times 1} \otimes \mathbf{b}'), \tag{3.12}$$

with $\mathbf{1}_{r\times 1}$ being a vector of ones of length r, $\mathbf{b} = (b_1, \ldots, b_p)'$ and $\boldsymbol{\Omega}$ denoting a covariance matrix with elements $\boldsymbol{\Omega}(A', A'')$, $\forall A', A'' \in \vec{A}$.

Summarizing all equations and allowing for d additional covariates in matrix $\mathbf{X}_2$, the final model becomes

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\text{vec}\big(\mathbf{B}(\vec{A})\big) + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \\ \text{vec}\big(\mathbf{B}(\vec{A})\big)\big|\tau, \mathbf{b}, \boldsymbol{\Phi}, \sigma^2 &\sim \mathcal{N}\big(\text{vec}(\bar{\mathbf{B}}), \sigma^2\boldsymbol{\Phi} \otimes \boldsymbol{\Omega}\big), \\ \boldsymbol{\varepsilon}|\sigma^2 &\sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}). \end{aligned} \tag{3.13}$$

To complete the model, we specify the following priors:

$$\begin{aligned} \boldsymbol{\beta}_2|\sigma^2 &\sim \mathcal{N}(\mathbf{0}, u \cdot \sigma^2\mathbf{I}), \\ \mathbf{b}|\sigma^2 &\sim \mathcal{N}(\mathbf{0}, u \cdot \sigma^2\mathbf{I}), \\ \boldsymbol{\Phi} &\sim \mathcal{W}^{-1}(\mathbf{P}, \nu), \end{aligned} \tag{3.14}$$

where u is a scalar that allows us to set the prior uncertainty on $\boldsymbol{\beta}_2$ and $\mathbf{b}$. Here, $\mathcal{W}^{-1}$ indicates the inverse Wishart distribution. We use a diffuse prior on σ^2, and the settings for the prior on τ will be discussed separately in Section 3.4.4 and provided for each illustration in the results section later in the chapter.

Note that to introduce the model, we considered the linear regression setup in Equation (3.9). However, models other than the linear Scheffé models and models for dependent variables that are not continuous can be developed in a similar manner. In Section 3.4.3, we work out details for a model in which the dependent variable is binary.

3.3.2 Variance-covariance structure of the mixture parameters

In this section, we exploit the structure of the variance-covariance matrix of the Gaussian process ($\sigma^2\boldsymbol{\Phi} \otimes \boldsymbol{\Omega}$) to model the correlation across the mixture parameters.

Consider again the mixture parameters stacked in the matrix $\mathbf{B}(\vec{A})$ as in Equation (3.8). One parameter $\beta_m(A_i)$ specifies the impact of a particular mixture proportion or a cross-product of proportions on the response at a specific value A_i of the amount variable. These parameters are not independent. First, our model incorporates the idea that for amount values that are close to each other, the model parameters are expected to be rather similar. Intuitively, the value of $\beta_m(A')$ should be similar to that of $\beta_m(A'')$ if $A' \approx A''$. In the model, this is captured by the correlation between the parameters $\beta_m(A')$ and $\beta_m(A'')$. The

correlation increases when the amounts A' and A'' are closer together. Second, at a given amount A', the parameters $\beta_1(A'), \beta_2(A'), \ldots, \beta_p(A')$ might also exhibit some correlation. For example, the effects of radio advertising and TV advertising may move up and down together as the advertising intensity changes. We allow for this type of correlation by letting $\beta_1(A'), \beta_2(A'), \ldots, \beta_p(A')$ be correlated as well.

We begin by specifying the correlation structure across different amount values for a given parameter β_m. We specify the elements of $\mathrm{Var}\big(\beta_m(\vec{A})\big) = \mathbf{\Omega}$ as

$$\mathbf{\Omega}(A', A'') = \exp\Big(-\frac{1}{2\tau^2}||A' - A''||^2\Big), \quad \tau > 0, \tag{3.15}$$

where A' and A'' are two amount values and τ denotes a model parameter. The function in Equation (3.15) is called the squared exponential (Gaussian) kernel. For any pair of amounts, A' and A'', a Gaussian process with this correlation function implies:

- $\beta_m(A')$ and $\beta_m(A'')$ will tend to have high correlation if A' and A'' are "close" to each another, since $||A' - A''||$ will then approach zero and $\mathbf{\Omega}(A', A'') = \exp\big(-\frac{1}{2\tau^2}||A' - A''||^2\big)$ will tend to one,
- $\beta_m(A')$ and $\beta_m(A'')$ will tend to have low correlation if A' and A'' are "far" apart, since $||A' - A''||$ will then be a large positive value and $\mathbf{\Omega}(A', A'') = \exp\big(-\frac{1}{2\tau^2}||A' - A''||^2\big)$ will tend to zero.

In other words, functions drawn from a Gaussian process with the Gaussian kernel will be locally smooth with high probability. This means that the mixture parameter values for amounts that are similar will also be similar. The similarity between the mixture parameters will decrease with the distance between A' and A''.

The parameter τ in Equation (3.15) controls the smoothness of the function of $\beta_m(A)$ as it determines how quickly $\beta_m(A)$ varies with A. By varying τ, we can in fact capture many different scenarios. By setting this parameter to zero or to a large positive value, we obtain standard models as special cases. In particular, if we take $\tau \to 0$, we allow separate mixture parameters for each value of the amount, as, if $A' \neq A''$ and τ approaches zero, $\mathbf{\Omega}(A', A'') = \exp(-\frac{1}{2\tau^2}||A' - A''||^2)$ tends to zero as well. On the other hand, if we let $\tau \to \infty$, we allow constant mixture parameters (i.e., independent of A), as, when τ increases, $\mathbf{\Omega}(A', A'') = \exp(-\frac{1}{2\tau^2}||A' - A''||^2)$ tends to one. By taking τ values between

0 and ∞, we can describe settings in between constant and separate mixture parameters, without restricting ourselves to any particular function for $\beta_m(A)$.

Instead of using some parametric function to model the mixture parameters in terms of the total amount variable, we only assume that the mixture parameters vary smoothly with the amount. This smoothness is controlled by a single parameter τ that can even be estimated. Such a specification is flexible enough to represent many different parametric forms that would require a very large number of parameters if formulated in a conventional way.

Whereas the correlation matrix $\mathbf{\Omega}$ captures the correlation structure for a given parameter β_m across different amount values, the covariance across the mixture parameters at a given amount is described by matrix $\mathbf{\Phi}$. At a given amount A', we have $\mathrm{Var}\big(\boldsymbol{\beta}(A')\big) = \mathbf{\Phi}$, where $\boldsymbol{\beta}(A') = \big(\beta_1(A'), \ldots, \beta_p(A')\big)'$ is the vector of the mixture parameters at A'. In the context of mixtures, this covariance is expected to be non-zero as a result of the direct impact of the total amount on the response. As an extreme illustration, consider a case where there are only two mixture ingredients, x_1 and x_2. Assume that their proportions have a constant impact on the response, while the amount A has a direct impact. Our proposed model will capture such a case with $\mathbf{\Phi}$ implying equal variances and a perfect correlation between $\beta_1(A)$ and $\beta_2(A)$. To see this, start with the model

$$y = \beta_1(A)x_1 + \beta_2(A)x_2 + \varepsilon.$$

The perfect correlation in combination with equal variances implies that we can write $\beta_1(A) = b_1 + \alpha(A)$ and $\beta_2(A) = b_2 + \alpha(A)$, where $\alpha(A)$ is a one-dimensional Gaussian process with mean zero and b_1 and b_2 are the means of $\beta_1(A)$ and $\beta_2(A)$, respectively. Using the mixture constraint, we can now rewrite the model as

$$\begin{aligned} y &= b_1x_1 + \alpha(A)x_1 + b_2x_2 + \alpha(A)x_2 + \varepsilon \\ &= \alpha(A) + b_1x_1 + b_2x_2 + \varepsilon. \end{aligned}$$

As in practice we usually expect a direct effect of the amount, $\mathbf{\Phi}$ will usually involve non-zero correlations. Note that, as $\mathbf{\Phi}$ is an unrestricted variance-covariance matrix, we need to restrict $\mathbf{\Omega}$ to be a correlation matrix to ensure identification.

3.4 Estimation

In this section, we discuss the Bayesian estimation of the model parameters in Equation (3.13) using Markov Chain Monte Carlo (MCMC) sampling. This approach requires taking draws from the joint posterior distribution of the model parameters (Gelman et al., 2013; Bishop, 2006; Greenberg, 2014; Zellner, 1996). However, sampling from the joint posterior density $p\Big(\tau, \mathbf{b}, \mathbf{B}(\vec{A}), \boldsymbol{\beta}_2, \boldsymbol{\Phi}, \sigma^2 \Big| \mathbf{y}\Big)$ is not directly feasible. Instead, we employ a Gibbs sampler (Casella and George, 1992) and repeatedly sample from conditional posterior distributions.

3.4.1 Sampling strategy

A straightforward Gibbs sampler where each parameter is drawn from its full conditional posterior is not efficient. The main reason for this is that we expect $\mathbf{B}(\vec{A})$ and τ to be strongly correlated: when τ is large (small) we expect quite similar (different) parameter values across the amount levels and vice versa. At the same time, $\mathbf{B}(\vec{A})$ and $\mathbf{b}$ are also likely to be strongly correlated. To reduce the dependence between the draws of $\mathbf{B}(\vec{A})$ and τ (and $\mathbf{b}$), we use the decomposition

$$p\Big(\tau, \mathbf{b}, \mathbf{B}(\vec{A}), \boldsymbol{\beta}_2 \Big| \mathbf{y}, \boldsymbol{\Phi}, \sigma^2\Big) = p\Big(\mathbf{b}, \mathbf{B}(\vec{A}), \boldsymbol{\beta}_2 \Big| \mathbf{y}, \boldsymbol{\Phi}, \sigma^2, \tau\Big) \times p\Big(\tau \Big| \mathbf{y}, \boldsymbol{\Phi}, \sigma^2\Big)$$

and apply Gibbs sampling steps for the latter distributions, where the sampling distribution of τ is not conditional on $\mathbf{B}(\vec{A})$ and $\mathbf{b}$. A Metropolis-Hastings (Chib and Greenberg, 1995) step within a Gibbs sampler is needed to sample τ. As a result, we iteratively sample from the four conditional distributions given at the right-hand side of Figure 3.1. Figure 3.1 graphically demonstrates how the sampling from the full posterior distribution is decomposed into iterative sampling from conditional posterior distributions.

In theory, one could treat the Gaussian process as a standard prior on $\mathrm{vec}\big(\mathbf{B}(\vec{A})\big)$. In our case, such an approach is numerically infeasible as some individual parameters may exhibit extreme correlations. This is especially true if some observed amount values are almost the same or if $\tau \to \infty$. The correlation matrix $\boldsymbol{\Omega}$ then becomes (nearly) singular and the traditional inverse of $\boldsymbol{\Omega}$ does not exist.

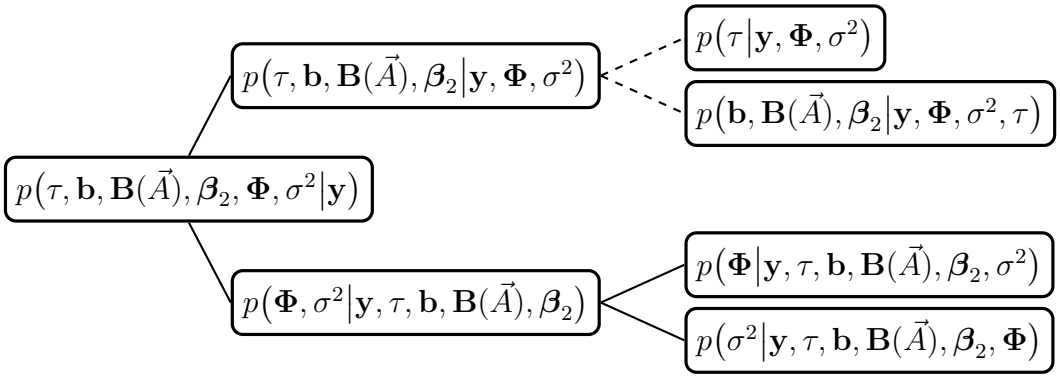


Figure 3.1: Decomposition of sampling from the joint posterior distribution into sampling from conditional posterior distributions used in the MCMC sampling. Dashed lines symbolize exact decompositions, solid lines symbolize decompositions based on Gibbs sampling

To make the estimation numerically tractable when $\tau \to \infty$ or some observed amount values are nearly identical, we use the singular value decomposition of the correlation matrix $\boldsymbol{\Omega}$, that is,

$$\boldsymbol{\Omega} = \mathbf{USV}' = \mathbf{USU}'. \tag{3.16}$$

Here, $\mathbf{U}$ is a real unitary matrix ($\mathbf{UU}' = \mathbf{I}$ with $\mathbf{I}$ an identity matrix) and $\mathbf{S}$ is a diagonal matrix. If $\boldsymbol{\Omega}$ is singular, some diagonal elements of $\mathbf{S}$ will be equal to zero. If it is nearly singular, these values will be close to zero. To improve the numerical stability of our estimation procedure, we replace these small diagonal elements by zeros. The threshold that we use to define a non-zero element is 10^{-6}. The corresponding matrix is denoted by $\mathbf{S}^*$. We now have $\boldsymbol{\Omega} \approx \mathbf{US}^*\mathbf{U}'$. Using this relation, we define the inverse of $\boldsymbol{\Omega}$ as

$$\boldsymbol{\Omega}^{-1} = \mathbf{US}^{*-1}\mathbf{U}', \tag{3.17}$$

where $\mathbf{S}^{*-1}$ is a diagonal matrix containing the reciprocals of all r^* non-zero diagonal entries of $\mathbf{S}^*$. This procedure is known as taking the Moore-Penrose pseudoinverse (Ben-Israel and Greville, 2003).

We next define the Choleski decomposition of $\boldsymbol{\Omega}$ as

$$\boldsymbol{\Omega}^{\frac{1}{2}} = \mathbf{US}^{*\frac{1}{2}}_{r \times r^*}, \tag{3.18}$$

where ${\mathbf{S}^*}^{\frac{1}{2}}_{r\times r^*}$ is the $r \times r^*$-dimensional matrix obtained by taking the square root of the non-zero elements of $\mathbf{S}^*$ and dropping the final $(r - r^*)$ zero columns. Note that $\boldsymbol{\Omega}^{\frac{1}{2}}{\boldsymbol{\Omega}^{\frac{1}{2}}}'$ indeed approximately equals $\boldsymbol{\Omega}$. The fact that $\boldsymbol{\Omega}$ is singular is reflected by the fact that the matrix $\boldsymbol{\Omega}^{\frac{1}{2}}$ is not square if $r^* \neq r$.

We exploit the singular value decomposition to sample $\mathrm{vec}\big(\mathbf{B}(\vec{A})\big)$ in the appropriate lower dimensional space if $\boldsymbol{\Omega}$ is (nearly) singular. We define $\mathrm{vec}\big(\mathbf{B}(\vec{A})\big) = \mathrm{vec}\big(\bar{\mathbf{B}}\big) + \big(\mathbf{F} \otimes \boldsymbol{\Omega}^{\frac{1}{2}}\big)\boldsymbol{\gamma}\big(\vec{A}\big)$ with $\boldsymbol{\gamma}(\vec{A}) \sim \mathcal{N}(\mathbf{0}_{pr^*\times 1}, \sigma^2\mathbf{I}_{pr^*\times pr^*})$, where $\mathbf{F}$ is the lower-triangular matrix resulting from the Choleski decomposition of $\boldsymbol{\Phi}$, i.e., $\boldsymbol{\Phi} = \mathbf{F}\mathbf{F}'$. Note that the distribution of $\boldsymbol{\gamma}(\vec{A})$ is a distribution in a lower dimensional space which naturally leads to $\mathrm{vec}\big(\mathbf{B}(\vec{A})\big) \sim \mathcal{N}\big(\mathrm{vec}\big(\bar{\mathbf{B}}\big), \sigma^2\boldsymbol{\Phi} \otimes \boldsymbol{\Omega}\big)$ as the implied distribution of $\mathrm{vec}\big(\mathbf{B}(\vec{A})\big)$, just as we defined before. Using this and the definition of $\bar{\mathbf{B}}$ (see Equation (3.12)) we can write

$$\begin{aligned}
\mathbf{X}\mathrm{vec}\big(\mathbf{B}(\vec{A})\big) &= \mathbf{X}\left(\mathrm{vec}\big(\bar{\mathbf{B}}\big) + \big(\mathbf{F} \otimes \boldsymbol{\Omega}^{\frac{1}{2}}\big)\boldsymbol{\gamma}\big(\vec{A}\big)\right) \\
&= \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{pmatrix} \mathbf{b} + \mathbf{X}\big(\mathbf{F} \otimes \boldsymbol{\Omega}^{\frac{1}{2}}\big)\boldsymbol{\gamma}\big(\vec{A}\big) \\
&= \mathbf{Z}\mathbf{b} + \mathbf{X}\big(\mathbf{F} \otimes \boldsymbol{\Omega}^{\frac{1}{2}}\big)\boldsymbol{\gamma}\big(\vec{A}\big),
\end{aligned}$$

where the matrix $\mathbf{Z} = (\mathbf{x}_1', \ldots, \mathbf{x}_N')'$ contains all explanatory variables in the Scheffé model used, ignoring the fact that different observations correspond to different amounts. Note that $\boldsymbol{\gamma}\big(\vec{A}\big)$ has a lower dimensionality, namely pr^*, than $\mathbf{B}(\vec{A})$, namely pr.

We can now sample $\boldsymbol{\gamma}(\vec{A})$ instead of $\mathrm{vec}\big(\mathbf{B}(\vec{A})\big)$ and avoid numerical issues due to potentially strong correlations. When some of the correlations in $\boldsymbol{\Omega}$ become too large, the singular value decomposition makes sure that the parameters are sampled in the lower dimensional space. Applying the above, we rewrite the model in Equation (3.13) and the priors in Equation (3.14) as

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}\mathrm{vec}\big(\mathbf{B}(\vec{A})\big) + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \\
&= \mathbf{X}\big(\mathbf{F} \otimes \boldsymbol{\Omega}^{\frac{1}{2}}\big)\boldsymbol{\gamma}\big(\vec{A}\big) + \mathbf{Z}\mathbf{b} + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} = \mathbf{X}^*\boldsymbol{\beta}^* + \boldsymbol{\varepsilon},
\end{aligned} \tag{3.19}$$

where $\mathbf{X}^* = \Big(\mathbf{X}\big(\mathbf{F} \otimes \mathbf{\Omega}^{\frac{1}{2}}\big) \;\; \mathbf{Z} \;\; \mathbf{X}_2\Big)$ and

$$\boldsymbol{\beta}^* = \begin{pmatrix} \boldsymbol{\gamma}(\vec{A}) \\ \mathbf{b} \\ \boldsymbol{\beta}_2 \end{pmatrix} \;\sim\; \mathcal{N}\left(\mathbf{0}_{p(r^*+1)+d\times 1}, \sigma^2 \begin{pmatrix} \mathbf{I}_{pr^*\times pr^*} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & u\mathbf{I}_{p\times p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & u\mathbf{I}_{d\times d} \end{pmatrix}\right)$$

or, more compactly,

$$\boldsymbol{\beta}^* \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{\Sigma}^*),$$

with $\mathbf{\Sigma}^* = \mathrm{diag}(\mathbf{I}_{pr^*\times pr^*}, u\mathbf{I}_{(p+d)\times(p+d)})$. We can next apply standard results to derive all the sampling steps needed for our model.

3.4.2 Sampling distributions

To obtain the conditional posterior distribution of τ, $p(\tau|\mathbf{y}, \mathbf{\Phi}, \sigma^2)$, we need to integrate over the distribution of $\boldsymbol{\gamma}(\vec{A})$, $\mathbf{b}$ and $\boldsymbol{\beta}_2$. The posterior distribution of τ is obtained as $p(\tau|\mathbf{y}, \mathbf{\Phi}, \sigma^2) \propto p(\mathbf{y}|\tau, \mathbf{\Phi}, \sigma^2)p(\tau)$, where $p(\tau)$ is the prior distribution of τ and

$$\begin{aligned} p\Big(\mathbf{y}\Big|\tau, \mathbf{\Phi}, \sigma^2\Big) &= \int_{\boldsymbol{\gamma}(\vec{A}),\mathbf{b},\boldsymbol{\beta}_2} p\Big(\mathbf{y}\Big|\tau, \mathbf{b}, \boldsymbol{\gamma}(\vec{A}), \boldsymbol{\beta}_2, \mathbf{\Phi}, \sigma^2\Big) p\Big(\boldsymbol{\gamma}(\vec{A}), \mathbf{b}, \boldsymbol{\beta}_2\Big|\tau, \mathbf{\Phi}, \sigma^2\Big) d\boldsymbol{\gamma}(\vec{A}) d\mathbf{b} d\boldsymbol{\beta}_2 \\ &= \int_{\boldsymbol{\beta}^*} p\Big(\mathbf{y}\Big|\tau, \boldsymbol{\beta}^*, \mathbf{\Phi}, \sigma^2\Big) p\Big(\boldsymbol{\beta}^*\Big|\sigma^2\Big) d\boldsymbol{\beta}^* \\ &\propto \exp\Big(-\frac{1}{2\sigma^2}\Big(\mathbf{w} - \mathbf{V}\hat{\beta}\Big)'\Big(\mathbf{w} - \mathbf{V}\hat{\beta}\Big)\Big) \times \Big|\Big(\mathbf{V}'\mathbf{V}\Big)^{-1}\Big|^{\frac{1}{2}}, \end{aligned} \tag{3.20}$$

where $\mathbf{w} = \big(\mathbf{y}' \;\; (\mathbf{0}_{p(r^*+1)+d\times 1})'\big)'$, $\mathbf{V} = \big(\mathbf{X}^{*\prime} \;\; (\mathbf{\Sigma}^{*-\frac{1}{2}})'\big)'$, $\hat{\beta} = (\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'\mathbf{w}$ and $\mathbf{\Sigma}^{*-1} = \mathbf{\Sigma}^{*-\frac{1}{2}\prime}\mathbf{\Sigma}^{*-\frac{1}{2}}$. The last step in this derivation follows from standard results for the linear model with a Gaussian prior applied to Equation (3.19).

Since the resulting posterior for τ, $p(\tau|\mathbf{y}, \mathbf{\Phi}, \sigma^2)$, is not of a known type, we apply the random-walk Metropolis-Hastings sampler to sample from it. We set the candidate generating function to be

$$\begin{aligned} \log\big(\tau^{\mathrm{Cand}}\big) &= \log\big(\tau^{\mathrm{Prev}}\big) + \eta, \\ \eta &\sim \mathcal{N}(0, \kappa^2). \end{aligned} \tag{3.21}$$

The acceptance probability of τ^{Cand} is then calculated as

$$\alpha = \min\left(\frac{p\big(\tau^{\text{Cand}}\big|\mathbf{y}, \mathbf{\Phi}, \sigma^2\big) g\big(\tau^{\text{Prev}}\big|\tau^{\text{Cand}}\big)}{p\big(\tau^{\text{Prev}}\big|\mathbf{y}, \mathbf{\Phi}, \sigma^2\big) g\big(\tau^{\text{Cand}}\big|\tau^{\text{Prev}}\big)}, 1\right), \tag{3.22}$$

where $g(\cdot)$ is the density of the candidate generating function in Equation (3.21) (Chib and Greenberg, 1995).

To sample $\boldsymbol{\beta}^*$, we consider the model given in Equation (3.19). The kernel of the conditional posterior distribution for $\boldsymbol{\beta}^*$ is

$$\boldsymbol{\beta}^*\Big|\mathbf{y}, \mathbf{\Phi}, \sigma^2, \tau \propto \exp\Big(-\frac{1}{2\sigma^2}\big(\boldsymbol{\beta}^* - \bar{\boldsymbol{\beta}}^*\big)'\big(\mathbf{X}^{*'}\mathbf{X}^* + \mathbf{\Sigma}^{*-1}\big)\big(\boldsymbol{\beta}^* - \bar{\boldsymbol{\beta}}^*\big)\Big), \tag{3.23}$$

where $\bar{\boldsymbol{\beta}}^* = \big(\mathbf{X}^{*'}\mathbf{X}^* + \mathbf{\Sigma}^{*-1}\big)^{-1}\mathbf{X}^{*'}\mathbf{y}$. This is the kernel of a multivariate normal distribution with mean $\bar{\boldsymbol{\beta}}^*$ and variance-covariance matrix $\sigma^2\big(\mathbf{X}^{*'}\mathbf{X}^* + \mathbf{\Sigma}^{*-1}\big)^{-1}$. Since $\bar{\mathbf{B}} = (\mathbf{1}_{r\times 1} \otimes \mathbf{b}')$, $\boldsymbol{\beta}^* = \big(\boldsymbol{\gamma}(\vec{A})' \;\; \mathbf{b}' \;\; \boldsymbol{\beta}_2'\big)'$ and $\text{vec}\big(\mathbf{B}(\vec{A})\big) = \text{vec}\big(\bar{\mathbf{B}}\big) + \big(\mathbf{F} \otimes \mathbf{\Omega}^{\frac{1}{2}}\big)\boldsymbol{\gamma}\big(\vec{A}\big)$, we can obtain draws for $\text{vec}\big(\mathbf{B}(\vec{A})\big)$ from draws for $\mathbf{b}$ and $\boldsymbol{\gamma}(\vec{A})$, see the discussion above.

We sample $\mathbf{\Phi}$ from the inverted Wishart distribution with parameters $\sigma^{-2}\big(\mathbf{B}(\vec{A}) - \bar{\mathbf{B}}\big)'\mathbf{\Omega}^{-1}$ $\big(\mathbf{B}(\vec{A}) - \bar{\mathbf{B}}\big) + \mathbf{P}$ and $r^* + \nu$, where $\mathbf{P}$ and ν give the prior scale and degrees of freedom, respectively.

Finally, we sample σ^2 from the inverted Gamma-2 distribution with parameter $(\mathbf{y} - \mathbf{X}^*\boldsymbol{\beta}^*)'(\mathbf{y} - \mathbf{X}^*\boldsymbol{\beta}^*) + \sigma^{-2}\boldsymbol{\beta}^{*\prime}\mathbf{\Sigma}^{*-1}\boldsymbol{\beta}^*$ and $N + p(r^* + 1) + d$ degrees of freedom.

3.4.3 Limited dependent variables

The ideas above can be easily generalized to deal with limited dependent variables. When the dependent variable $\mathbf{y}$ is binary, we employ the estimation procedure by Albert and Chib (1993); McCulloch and Rossi (1994, 2000); Allenby and Rossi (1999); Train (2009). Write the model as

$$y_i = \begin{cases} 1 \text{ if} & z_i = X_i\text{vec}\big(\mathbf{B}(\vec{A})\big) + X_{2i}\boldsymbol{\beta}_2 + \varepsilon_i > 0, \\ 0 \text{ if} & z_i = X_i\text{vec}\big(\mathbf{B}(\vec{A})\big) + X_{2i}\boldsymbol{\beta}_2 + \varepsilon_i \leq 0, \end{cases}$$

for $i = 1, \ldots, N$, where $\varepsilon_i \sim \mathcal{N}(0, 1)$ and all other details of the model stay the same. The only detail to note is that the variance of ε_i is restricted to one. Therefore, in the notation of the previous sections, we restrict σ^2 to be one. For parameter inference, we sample z_i,

$i = 1, \ldots, N$, alongside the other parameters as part of the Gibbs sampler (keeping $\sigma^2 = 1$ fixed). Denote $\mathbf{z} = (z_1, \ldots, z_N)'$. The conditional distribution for the only additional step is given by

$$p\big(\mathbf{z}\big|\mathbf{y}, \tau, \mathbf{b}, \mathrm{vec}\big(\mathbf{B}(\vec{A})\big), \boldsymbol{\beta}_2, \boldsymbol{\Phi}, \sigma^2 = 1\big).$$

All elements of $\mathbf{z}$ are independent of each other conditional on the model parameters. Hence, for the i^{th} element z_i, the distribution reduces to

$$z_i\big|y_i, \mathrm{vec}\big(\mathbf{B}(\vec{A})\big), \boldsymbol{\beta}_2 \sim \begin{cases} \mathcal{N}\Big(X_i\mathrm{vec}\big(\mathbf{B}(\vec{A})\big) + X_{2i}\boldsymbol{\beta}_2, 1\Big)\mathcal{I}\big(z_i > 0\big) & \text{if } y_i = 1, \\ \mathcal{N}\Big(X_i\mathrm{vec}\big(\mathbf{B}(\vec{A})\big) + X_{2i}\boldsymbol{\beta}_2, 1\Big)\mathcal{I}\big(z_i \leq 0\big) & \text{if } y_i = 0, \end{cases}$$

with $\mathcal{I}(\cdot)$ denoting the indicator function and $i = 1, \ldots, N$.

3.4.4 Prior specification for τ

Some care is needed when choosing a prior distribution for τ. In this section, we provide some intuition on how we specify this prior.

First, from Equation (3.15), the correlation between $\beta_m(A')$ and $\beta_m(A'')$ for $A' \neq A''$ depends only on the difference between A' and A''. Note that for larger amount values, a larger value of τ is required to represent the same level of correlation. The prior required for τ therefore depends on the scale of the amount variable. To deal with this issue, we standardize the total amount variable, by dividing it by its standard deviation. Note that standardization preserves the ranking and the pairwise ratios of the total amount values. Moreover, from the prior distribution for τ when the total amount variable is standardized, we can always derive the corresponding prior distribution for τ for the original amount values. To show this, denote the standard deviation of the amount variable in the data by S. Write then

$$\begin{aligned} \boldsymbol{\Omega}(A', A'') &= \exp\Big(-\frac{1}{2\tau^2}\big|\big|A' - A''\big|\big|^2\Big) \\ &= \exp\Big(-\frac{1}{2(\tau/S)^2}\big|\big|A'/S - A''/S\big|\big|^2\Big) = \exp\Big(-\frac{1}{2\tau^{*2}}\big|\big|A^{*'} - A^{*''}\big|\big|^2\Big), \end{aligned}$$

where $\tau^* = \tau/S$ and $A^* = A/S$ is the standardized amount value. Now, if we consider the standardized amounts, we begin by specifying a prior distribution for τ^*. Then, to obtain the

corresponding distribution for τ for the original amount values, we can use $\tau = \tau^* S$, where the distribution for τ^* is known.

Second, in some cases, we may want to use a prior distribution with a strictly positive domain, that is, a prior that sets zero probability on no correlation. Such a prior is especially useful for data where only one observation per amount value is available. Here, if we allowed the Gaussian process to have zero correlation, we would end up with independent parameters for each individual observation. Naturally, such a specification does not make sense. Even if multiple observations per amount level are available, one may still want to impose such a prior if one expects the amount levels to be somehow related to some unobserved factors in the data generating process.

Finally, when choosing a prior for τ we are in fact specifying a prior on the correlation structure across different amount values. An uninformative prior for τ may sometimes lead to a very informative specification for the correlations. Therefore, after specifying a prior for τ, it is useful to inspect the implied prior for the correlations (see Gelman (2006); Gilmour and Goos (2009) for a related discussion).

3.5 Illustrations

In this section, we consider two data sets to illustrate our approach. The first data set describes how mice react to hormone mixtures administered at three different amount levels. The dependent variable here is continuous and describes cornification of the vaginal epithelium. Using this data set, we demonstrate that two common models in the mixture-amount literature are special cases of our model. In these special cases, a certain functional form is assumed for the mixture parameters.

Often, it is not a priori known how the mixture parameters depend on the total amount meaning that functional form assumptions may not be justified. Therefore, we demonstrate next how, in our approach, we estimate this relationship without making any parametric assumptions. We do so using a realistic data set, which describes to what extent women recognize advertisements run in magazines and/or on television with different intensities as measured by Gross Rating Points (GRPs) (operationalized as the amount variable, with 52 unique values). The dependent variable here is binary and indicates whether an advertising campaign is recognized (1) or not (0).

3.5.1 Mice experiment

For the first example, we consider data from Claringbold (1955) who presented an experiment involving 10 different mixtures of three distinct hormones administered to 10 groups of 12 mice each. Each hormone mixture was studied at three amount levels, 0.75×10^{-4} μg (A_1), 1.50×10^{-4} μg (A_2) and 3.00×10^{-4} μg (A_3), and so there were 30 experimental runs in total. The response variable of interest is the fraction of mice in each group (out of 12) that responded to each of the 30 mixture-amount combinations. The dependent variable considered is the angular transformation of the fractions, see Claringbold (1955) for details. We replicate the data in Table 3.1.

Hormone proportion			Percent response ($p \times 100$)			Angular response (y)		
x_1	x_2	x_3	A_1	A_2	A_3	A_1	A_2	A_3
1	0	0	17	42	83	24.09	40.20	65.91
0	1	0	58	58	100	49.80	49.80	81.70
0	0	1	25	50	42	30.00	45.00	40.20
$\frac{2}{3}$	$\frac{1}{3}$	0	0	33	75	8.30	35.26	60.00
$\frac{1}{3}$	$\frac{2}{3}$	0	33	33	75	35.26	35.26	60.00
$\frac{2}{3}$	0	$\frac{1}{3}$	0	25	75	8.30	30.00	60.00
$\frac{1}{3}$	0	$\frac{2}{3}$	25	42	42	30.00	40.20	40.20
0	$\frac{2}{3}$	$\frac{1}{3}$	17	33	67	24.09	35.26	54.74
0	$\frac{1}{3}$	$\frac{2}{3}$	33	33	58	35.26	35.26	49.80
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	17	25	58	24.09	30.00	49.80

Table 3.1: Data for the mice experiment

We use these data to estimate the parameters of two simple mixture-amount models and to demonstrate that they are special cases of the methodology we introduce in this chapter. Consider first a simple linear Scheffé model for the complete dataset, ignoring the amount (see Equation (3.1)):

$$y_i = \beta_1^0 x_{1i} + \beta_2^0 x_{2i} + \beta_3^0 x_{3i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_0^2). \tag{3.24}$$

Next, consider the same linear regression for each observed amount level separately, that is,

$$y_i = \beta_1^1 x_{1i} + \beta_2^1 x_{2i} + \beta_3^1 x_{3i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_1^2), \quad \text{if } i \text{ corresponds to } A_1, \tag{3.25}$$

$$y_i = \beta_1^2 x_{1i} + \beta_2^2 x_{2i} + \beta_3^2 x_{3i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_2^2), \quad \text{if } i \text{ corresponds to } A_2, \tag{3.26}$$

$$y_i = \beta_1^3 x_{1i} + \beta_2^3 x_{2i} + \beta_3^3 x_{3i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_3^2), \quad \text{if } i \text{ corresponds to } A_3. \tag{3.27}$$

As priors, we take $\boldsymbol{\beta}^j = (\beta_1^j, \beta_2^j, \beta_3^j)' | \sigma_j^2 \sim \mathcal{N}(\mathbf{0}_{3\times 1}, 10^3 \times \sigma_j^2 \mathbf{I}_{3\times 3})$ for the mixture parameters ($j = 0, 1, 2, 3$ corresponding to the models in Equations (3.24)-(3.27), respectively) and the diffuse priors for the variance parameters. We display summary statistics for the posterior sample of $100,000$ draws in Table 3.2.

	Model			
	Ignoring amount (Eq. (3.24), $j = 0$)	Considering A_1 only (Eq. (3.25), $j = 1$)	Considering A_2 only (Eq. (3.26), $j = 2$)	Considering A_3 only (Eq. (3.27), $j = 3$)
β_1^j	35.66	12.03	33.68	61.31
	(6.35)	(6.75)	(4.43)	(4.85)
β_2^j	50.49	40.24	40.58	70.59
	(6.34)	(6.75)	(4.44)	(4.87)
β_3^j	34.60	28.47	38.59	36.78
	(6.35)	(6.75)	(4.47)	(4.84)
σ_j^2	241.33	91.54	39.61	47.29
	(67.07)	(53.25)	(22.77)	(27.36)

Table 3.2: Posterior means and standard deviations (in parentheses) for the parameters of the models in Equations (3.24)-(3.27)

The model in Equation (3.24) assumes the same parameters irrespective of the amount value. The models in Equations (3.25)-(3.27) assume different, and unrelated, parameters for each amount value. We argue that these two cases are nested within the model which we introduce in this chapter. As a result, the model based on the Gaussian process prior should be able to replicate the results obtained above. Prior to analysis, to make τ less dependent on the scale of the amount variable (see Section 3.4.4), we standardize the amount. In order to obtain a model where the mixture parameters are independent of the amount, as in Equation (3.24), we set a large value for τ ($\tau = 10,000$) in Equation (3.15). Furthermore, we fix $\mathbf{b} = \mathbf{0}_{pr\times 1}$ and $\boldsymbol{\Phi} = 10^3 \times \mathbf{I}_{p\times p}$, to match the typical uninformative regression setting. In fact, then the posterior means of $\beta_1^j, \beta_2^j, \beta_3^j, \sigma_j^2$, $j = 0, 1, 2, 3$, will be plain OLS estimates.

The second column in Table 3.3 gives the summary statistics for the posterior sample of $100,000$ draws. We can clearly see that the estimates of the mixture parameters are indeed constant with respect to the amount variable. Furthermore, they are not very different from the parameter estimates obtained for the model in Equation (3.24), shown in the second column of Table 3.2.

	Model		
	constant parameters ($\tau = 10,000$)	different parameters per amount ($\tau = 0$)	benchmark models (Eq. (3.25) - (3.27))
$\beta_1(A_1)$	35.67	12.02	12.03
	(6.34)	(5.03)	(6.75)
$\beta_1(A_2)$	35.67	33.66	33.68
	(6.34)	(5.05)	(4.43)
$\beta_1(A_3)$	35.67	61.31	61.31
	(6.34)	(5.05)	(4.85)
$\beta_2(A_1)$	50.48	40.23	40.24
	(6.36)	(5.05)	(6.75)
$\beta_2(A_2)$	50.48	40.57	40.58
	(6.36)	(5.04)	(4.44)
$\beta_2(A_3)$	50.48	70.56	70.59
	(6.36)	(5.03)	(4.87)
$\beta_3(A_1)$	34.61	28.48	28.47
	(6.33)	(5.05)	(6.75)
$\beta_3(A_2)$	34.61	38.59	38.59
	(6.33)	(5.03)	(4.47)
$\beta_3(A_3)$	34.61	36.77	36.78
	(6.33)	(5.04)	(4.84)
σ^2	241.60	50.92	-
	(66.80)	(14.00)	-

Table 3.3: Posterior means and standard deviations (in parentheses) for parameters obtained from the Gaussian process prior model (columns 2-3) and from the benchmark models (column 4)

To obtain a model with separate independent parameters for each observed amount level, as in Equations (3.25)-(3.27), we set $\tau = 0$ in Equation (3.15) and fix $\mathbf{b} = \mathbf{0}_{pr\times 1}$ and $\mathbf{\Phi} = 10^3 \times \mathbf{I}_{p\times p}$. The posterior means and standard deviations (in parentheses) for the parameters are given in the third column of Table 3.3. For ease of comparison, we repeat the estimates from Table 3.2 (columns 3-5) in the last column of Table 3.3. The corresponding parameter estimates are again very similar. As a result, the mixture-amount model based on the Gaussian process prior indeed covers the two extreme scenarios described in Equations (3.24)-(3.27).

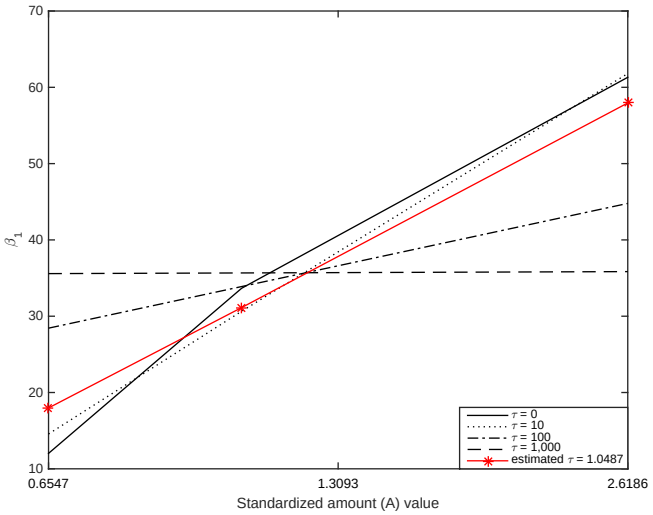
Using $\tau = 0$ and $\tau = 10,000$ led to either independent or constant parameters across amount values. By choosing $0 < \tau < \infty$, we can describe many different intermediate settings without explicitly assuming a particular parametric form for each $\beta_m(A)$. We illustrate this in Figure 3.2 where we use τ values of $0; 10; 100$ and $1,000$ and plot the corresponding posterior means of the mixture parameters with respect to the standardized amount variable. In the same figure, we plot the posterior means of the mixture parameters for an estimated $\tau = 1.0487$ (in red), obtained using the prior $p(\tau) \sim \ln\mathcal{N}(0.1, 0.4^2)$ and uninformative priors for the other parameters. It is clear that, by changing τ, we can describe many different scenarios and, when $\tau \to \infty$, the mixture parameters no longer change with the amount.

In this section, we demonstrated that our model based on the Gaussian process prior, if τ is chosen accordingly, can in fact replicate two simple models for mixture-amount data. We also showed the mixture parameters for an estimated τ value without getting into detail of how we do this. In the next section, we consider the second data set and demonstrate the estimation of τ together with β_m, $\mathbf{b}$, $\mathbf{\Phi}$ and σ^2.

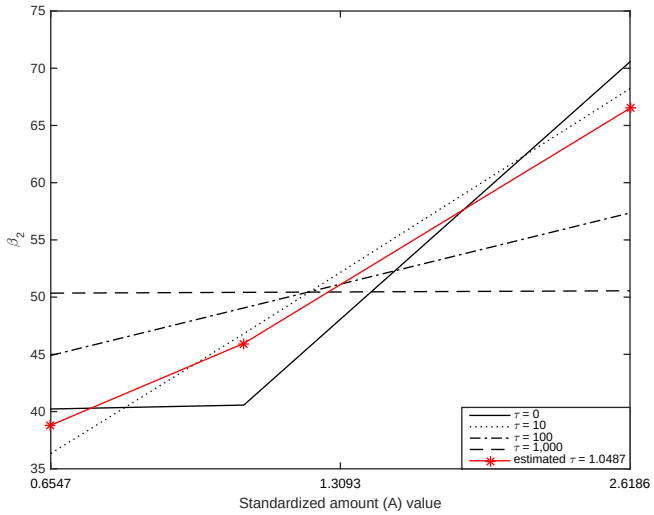
3.5.2 Advertising campaign recognition

In this section, we consider an application concerning advertising campaign recognition. In a questionnaire, individual female respondents indicated whether they recognized various skin and hair care advertising campaigns. The dependent variable takes the value 1 if a campaign is recognized and 0 otherwise. The mixture variables describing each campaign are proportions of the total advertising exposure (A) in magazine (x_1) and on television (TV) (x_2), which make up 100% for every campaign. There are differences in the advertising campaigns across regions, where these differences are in the total exposure to advertising as well as in the proportions across TV and magazines. For each respondent, we know in which of the regions she lives. As a measure of the advertising campaign exposure, we use Gross Rating Points (GRPs), where a GRP is defined as a percentage of the target audience reached by a campaign (De Pelsmacker et al., 2010).

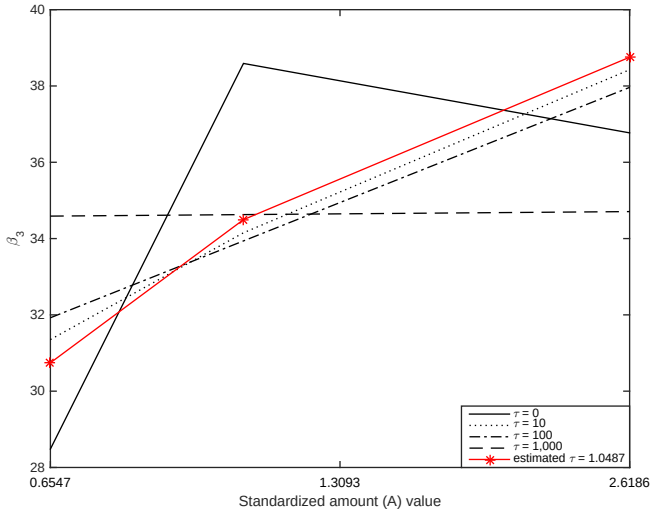
There are 52 advertising campaigns in the data set in total corresponding to 52 unique total amount values. We provide the histogram of the total amount values in our data set in Figure 3.3, where, for ease of comparison, we give both original and standardized amount values on the lower and upper axes, respectively. As we can see, the share of the ads with



(a) $\beta_1(\vec{A})$



(b) $\beta_2(\vec{A})$



(c) $\beta_3(\vec{A})$

Figure 3.2: Posterior means for the mixture parameters as a function of the standardized amount for different values of τ

less than 300 GRPs is the largest in our sample, whereas we have much fewer ads with more than 500 GRPs. From Figure 3.4, we can see the proportions of magazine advertising versus the total amount values (original amount values on the bottom axis and standardized amount values on the top axis). Notice that, in our data set, for large(r) total amount values, the magazine advertising tends to be low(er), and more of the total advertising exposure tends to

be invested in TV advertising. Furthermore, there are no observations with a proportion of magazine advertising between 40% and 99%.

The advertising campaigns ran in magazines and/or on TV in the period of June-December, 2011, in the Netherlands and two regions of Belgium, Flanders and Wallonia. For selected campaigns, consumer responses were recorded by means of an online survey at 5 different points in time (the so-called waves). There are approximately 500 respondents per wave and per region. In Flanders, campaigns comprise 4 brands, and there is a total of 9,490 individual observations. There are 4 brands in Wallonia (7,786 individual responses) and 6 brands in the Netherlands (9,509 individual responses). Note that, for some brands, there were multiple campaigns. In total, there are 26,785 responses from 6,679 respondents in our data set. We provide a subset of the data set in Table 3.4. Note that our data are somewhat restrictive as each campaign ran in a specific region, for one brand, at one time point only. Within every campaign, there is no variation in the media mix. More information about the data can be found in Aleksandrovs et al. (2015), who use the ads run in Belgium to introduce mixture-amount modeling in the advertising literature.

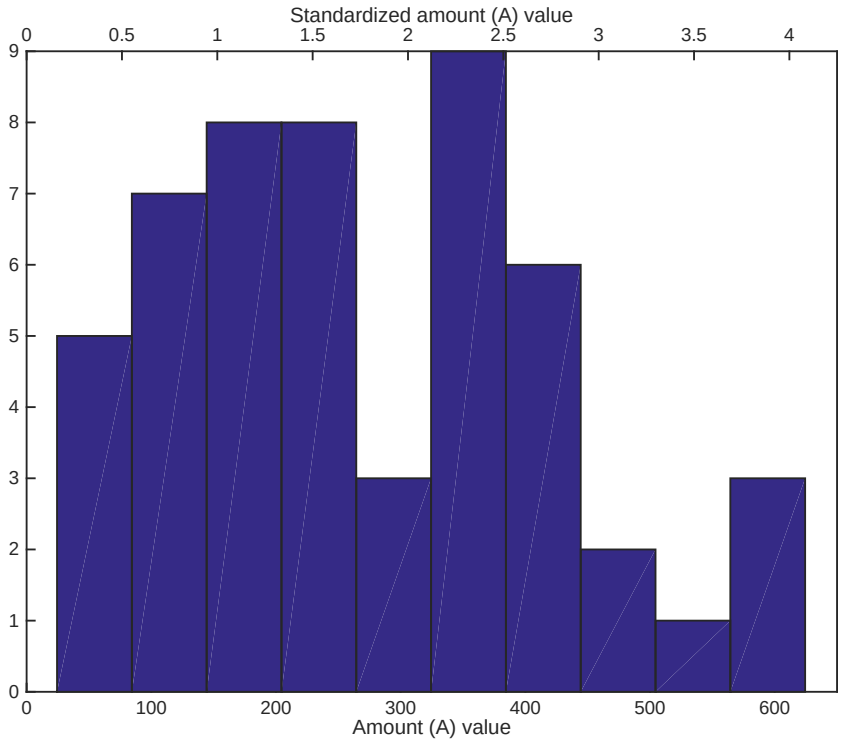


Figure 3.3: Histogram of the amount values in the advertising campaign data set

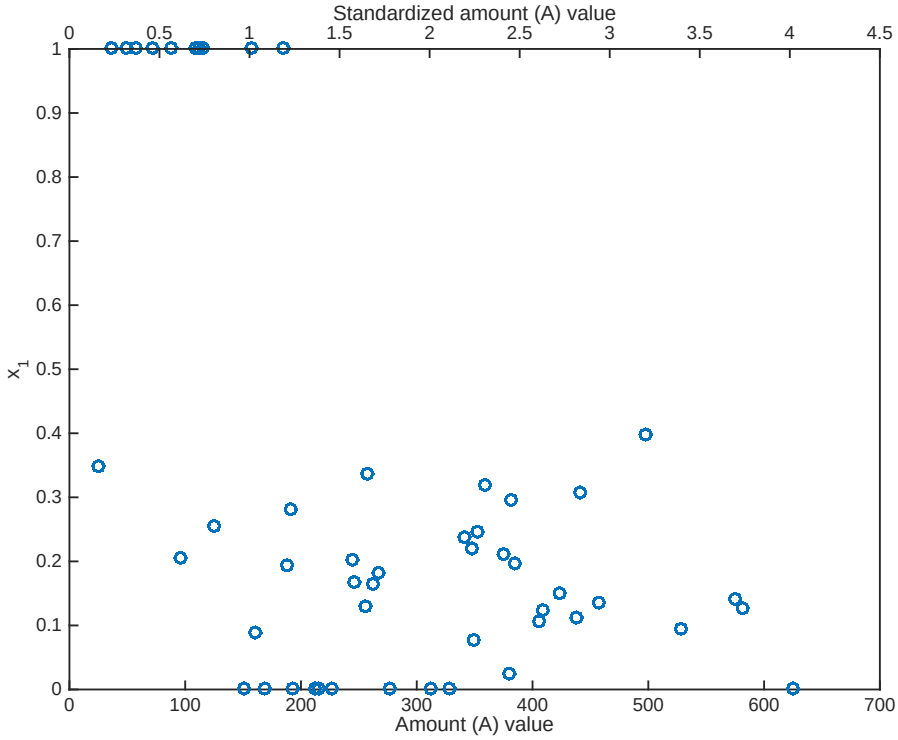


Figure 3.4: Scatter plot of the proportions of magazine advertising (x_1) and the total amount (A) values

Parameter estimation

In this section, we first consider the second-order Scheffé model for the mixture variables (see Equation (3.2)). As additional control variables, we include dummy variables for the region, brand and wave. The model for the latent variables driving the campaign recognition of the respondents can then be written as

$$\mathbf{z} = \mathbf{X}\text{vec}\big(\mathbf{B}(\vec{A})\big) + \mathbf{D}_{\text{Region}}\boldsymbol{\beta}_{\text{Region}} + \mathbf{D}_{\text{Brand}}\boldsymbol{\beta}_{\text{Brand}} + \mathbf{D}_{\text{Wave}}\boldsymbol{\beta}_{\text{Wave}} + \boldsymbol{\varepsilon}, \tag{3.28}$$

where we define $\mathbf{X}$ and $\text{vec}\big(\mathbf{B}(\vec{A})\big)$ as in Equation (3.11) with $\mathbf{x}_i = (x_{1i}, x_{2i}, x_{1i}x_{2i})'$ and $\mathbf{D}_{\text{Region}}$, $\mathbf{D}_{\text{Brand}}$, $\mathbf{D}_{\text{Wave}}$ are matrices with dummy coded columns which correspond to the observations' region, brand and wave, respectively, and $\boldsymbol{\beta}_{\text{Region}}$, $\boldsymbol{\beta}_{\text{Brand}}$ and $\boldsymbol{\beta}_{\text{Wave}}$ are the corresponding vectors of, what we call, non-mixture parameters. The reference region in the model is Wallonia. The reference brand is brand 6 and the reference wave is wave 5.

Following the discussion in Section 3.4.4, we standardize the amount variable and use $\mathcal{U}(0.75, 2)$ as a prior distribution for τ (note the lower bound that we impose on τ). We use

Campaign	Wave	Region	Brand	ID	GRP_{MAG} (x_1)	GRP_{TV} (x_2)	GRP	Recognition
1	1	Netherlands	2	4791	31.70 (0.25)	92.80 (0.75)	124.50	0
1	1	Netherlands	2	4796	31.70 (0.25)	92.80 (0.75)	124.50	0
1	1	Netherlands	2	4787	31.70 (0.25)	92.80 (0.75)	124.50	1
1	1	Netherlands	2	4810	31.70 (0.25)	92.80 (0.75)	124.50	1
2	1	Netherlands	2	4810	48.70 (0.18)	218.60 (0.82)	267.30	1
...	...	...	...	...	...	...		
12	2	Wallonia	2	2160	86.60 (0.34)	170.00 (0.66)	256.60	0
13	2	Wallonia	5	2160	135.00 (0.31)	305.80 (0.69)	440.80	1
...	...	...	...	...	...	...		
31	3	Netherlands	6	6530	108.10 (1.00)	0.00 (0.00)	108.10	0
...	...	...	...	...	...	...		
51	1	Flanders	4	35613	19.60 (0.21)	76.00 (0.79)	95.60	1
52	5	Flanders	4	6261	0.00 (0.00)	276.50 (1.00)	276.50	0

Table 3.4: Data for advertising campaign recognition

$\mathcal{W}^{-1}(3 \times \mathbf{I}_{3\times 3}, 7)$ as a prior distribution for $\mathbf{\Phi}$, $\mathcal{N}(\mathbf{0}_{11\times 1}, 10 \times \mathbf{I}_{11\times 11})$ as a prior distribution for $(\boldsymbol{\beta}'_{\text{Region}}\ \boldsymbol{\beta}'_{\text{Brand}}\ \boldsymbol{\beta}'_{\text{Wave}})'$ (that is, we use $u = 10$, see Section 3.4) and a similar distribution for $\mathbf{b}$. Since our dependent variable is a $0/1$ variable, we formulate the problem as a choice model and therefore set $\sigma^2 = 1$ (see Section 3.4.3). Note also that 7 degrees of freedom of the prior distribution of $\mathbf{\Phi}$ is the minimum required for the expected value and variance of $\mathbf{\Phi}$ to exist, and the prior of $\mathbf{\Phi}$ implies $\text{E}(\mathbf{\Phi}) = \mathbf{I}_{3\times 3}$, where $\mathbf{I}_{3\times 3}$ denotes an identity matrix.

As initial values, we use $\tau_{\text{init}} = 1$, $\mathbf{\Phi}_{\text{init}} = \mathbf{I}_{3\times 3}$, $(\boldsymbol{\beta}'_{\text{Region}}\ \boldsymbol{\beta}'_{\text{Brand}}\ \boldsymbol{\beta}'_{\text{Wave}})' = \mathbf{0}_{11\times 1}$ and $\mathbf{b} = \mathbf{0}_{3\times 1}$. We set κ in Equation (3.21) to 0.2, resulting in an acceptance rate of 42.09% in Equation (3.22), which is close to the suggested target in Robert and Casella (2010). We use $20,000$ iterations in the estimation and subsequently disregard $10,000$ samples as a burn-in. To show convergence, we plot the Markov chain for τ in Figure 3.5. The posterior mean, standard deviation and 95% HPD interval for τ are 0.87, 0.10 and $[0.75 \quad 1.07]$, respectively. To demonstrate the implied correlation structure for the mixture parameters across different amounts at the posterior mean of τ, we plot the correlation versus the differences in the standardized amount values in Figure 3.6. The circles denote the implied correlation values at the smallest observed distance (0.0026) and largest observed distance (3.9105) in our data set. In Table 3.5, we give the posterior estimates of $\boldsymbol{\beta}_{\text{Region}}$, $\boldsymbol{\beta}_{\text{Brand}}$ and $\boldsymbol{\beta}_{\text{Wave}}$. From Table 3.5 we see that, when controlling for an ad configuration, brand and wave, on average, women from Wallonia recognize a larger number of campaigns than their Dutch or Flemish counterparts. Further, a larger number of ads is recognized in wave 5 than in other waves, all other covariates in Equation (3.28) being equal. Finally, the posterior mean of $\mathbf{\Phi}$ is

$$\begin{pmatrix} 2.83 & -0.07 & -2.24 \\ -0.07 & 0.60 & 0.26 \\ -2.24 & 0.26 & 3.73 \end{pmatrix}.$$

$\boldsymbol{\beta_D}$	j	$\mathbb{E}(\boldsymbol{\beta_D}\|\mathbf{y})$	$\text{StDev}(\boldsymbol{\beta_D}\|\mathbf{y})$	95% HPD interval	
$\boldsymbol{\beta}_{\text{Region}_j}$	Flanders	-0.17	0.02	-0.22	-0.13
	Netherlands	-0.21	0.03	-0.27	-0.15
$\boldsymbol{\beta}_{\text{Brand}_j}$	1	0.04	0.04	-0.04	0.11
	2	-0.58	0.06	-0.69	-0.46
	3	-0.08	0.03	-0.15	-0.02
	4	0.29	0.03	0.23	0.35
	5	-0.07	0.08	-0.21	0.08
$\boldsymbol{\beta}_{\text{Wave}_j}$	1	-0.18	0.04	-0.25	-0.12
	2	-0.28	0.03	-0.34	-0.21
	3	-0.14	0.03	-0.21	-0.08
	4	-0.10	0.04	-0.18	-0.03

Table 3.5: Posterior means, standard deviations and 95% HPD intervals for the non-mixture parameters for the advertising campaign data set

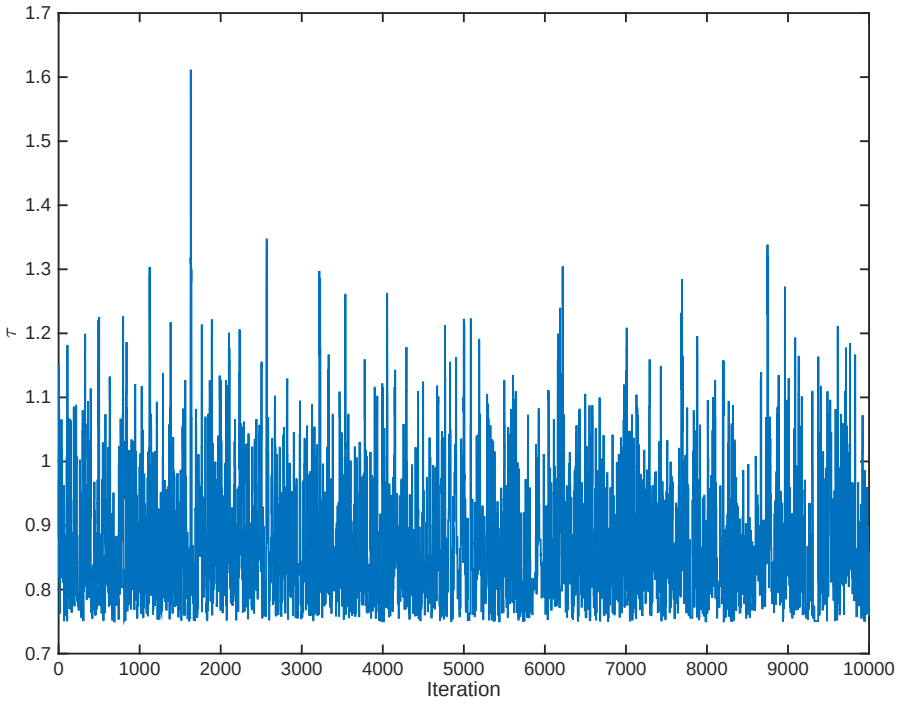


Figure 3.5: Posterior τ draws after a burn in of $10,000$ observations for the advertising campaign data set

In Figure 3.7, we plot the posterior means of $\beta_1(\vec{A})$, $\beta_2(\vec{A})$ and $\beta_3(\vec{A})$ (blue inner curves) together with 95% HPD intervals (green curves above and below) versus the standardized amount. The horizontal red lines correspond to the posterior means of b_1, b_2 and b_3. The dots on the curves denote the observed amount values in our data set. We see that the effects of both magazine and TV advertising and also the effect of the interaction of the two advertising media vary smoothly with respect to the total advertising exposure. Note that none of the shapes among $\beta_1(\vec{A})$, $\beta_2(\vec{A})$ and $\beta_3(\vec{A})$ is linear or quadratic, which are the functions commonly assumed in standard mixture-amount models in the literature. Furthermore, they are different for different mixture variables. Importantly, the values of $\beta_1(\vec{A})$, $\beta_2(\vec{A})$ and $\beta_3(\vec{A})$ are not constant and differ substantially from the means $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$, which demonstrates that the effect of mixture proportions is not constant with respect to the amount. As the data contain a larger number of campaigns with small GRPs (see Figure 3.3), the uncertainty depicted by the green curves corresponding to the 95% HPD intervals is the smallest for observations at lower GRPs. The lack of both campaigns with larger GRPs and explanatory variables which vary over the campaigns in the data set lead to somewhat wider HPD intervals, especially for the campaigns with larger GRPs.

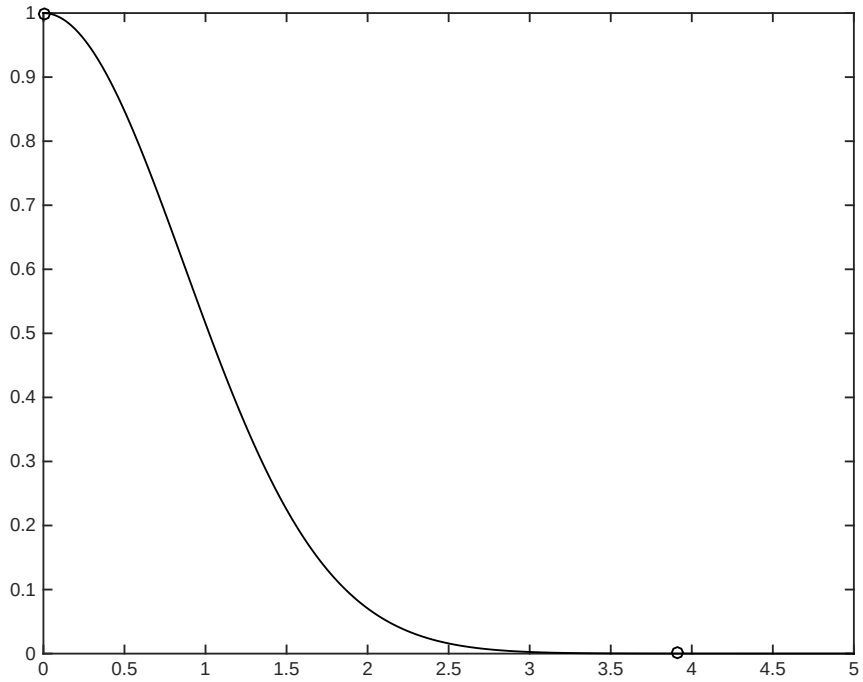
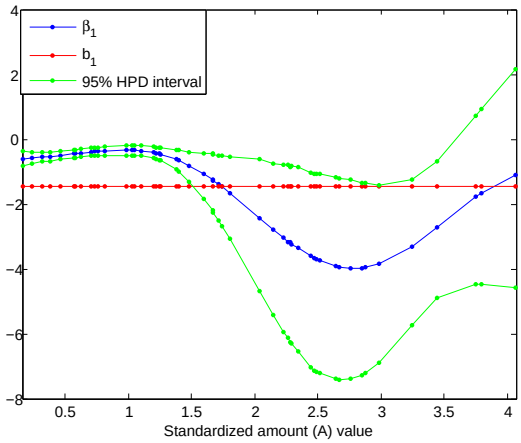
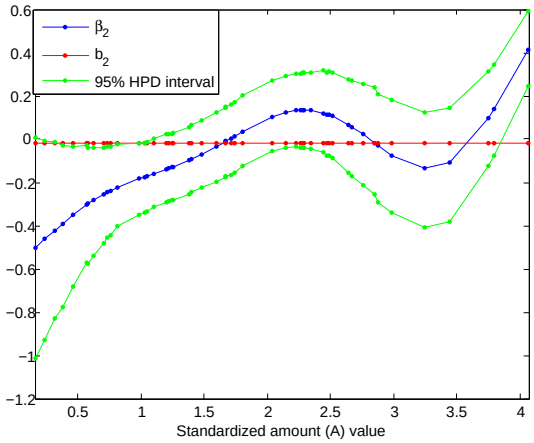


Figure 3.6: Implied correlation at the posterior mean of τ versus the difference in the standardized amount. The circles denote implied correlation at the smallest and largest observed differences in the advertising campaign data set

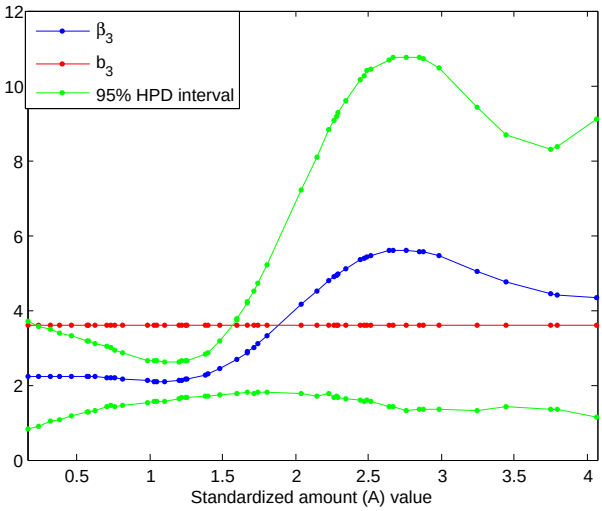
In Figure 3.8, using the estimated model in Equation (3.28), we demonstrate how the probability of the campaign recognition changes for different values of the total advertising exposure and the advertising media mix, at the average values for the region, brand and wave. Using the posterior distribution of the parameters, we calculate the probability of recognizing a campaign for values of the standardized amount ranging from 0.15 to 4.1 and the proportion of magazine advertising ranging from 0.1 to 1. The ranges of the amount and magazine proportion match the range in the observed data. To obtain the posterior distribution for the mixture parameters at amount values that are not used in the data, we use Equation (3.7). As expected, we see that the largest recognition probability is achieved for the largest total advertising exposure values. The large dip in the probability of recognition, at high proportions of magazine advertising and values of the standardized amount of about 3, can be explained by the following. First, as it can be seen from Figure 3.3, our data set does not contain many campaigns where the total advertising exposure is around 480 GRPs (standardized amount around 3). Furthermore, for such large(r) advertising exposure values, the observed proportion of magazine advertising is always low, see Figure 3.4. The result of this is that the estimation uncertainty around the recognition probabilities is quite large in



(a) $\beta_1(\vec{A})$ (Magazines)



(b) $\beta_2(\vec{A})$ (TV)



(c) $\beta_3(\vec{A})$ (Magazines$\times$TV)

Figure 3.7: Posterior means for the mixture parameters versus the standardized amount

this range. To avoid clutter, we do not show this estimation uncertainty in Figure 3.8. The dip itself is explained by the fact that, in the data set, the campaign recognition was very low at relatively similar observations. An interesting observation is that, when the total advertising exposure is low(er), to maximize the recognition, it seems to be wiser to invest more in magazine advertising. On the other hand, when the total advertising exposure is large, to maximize the campaign recognition, it is wiser to invest more in TV advertising. This information is important when choosing between advertising media for a given advertising exposure. Then, this type of graph lends itself to evaluating tradeoffs when deciding how much of an advertising budget to allocate in total and per medium.

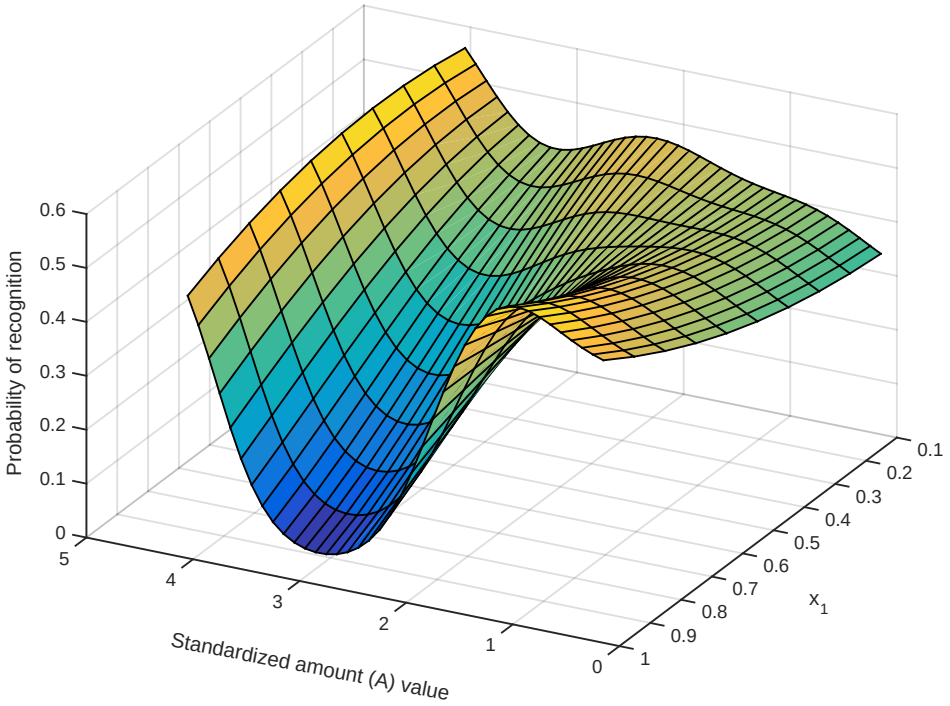


Figure 3.8: Campaign recognition probabilities for different values of the standardized amount and the proportion invested in magazine advertising (x_1) for the model in Equation (3.28)

Forecasting performance

In this section, we demonstrate the performance of our model in terms of forecasting recognition for amount values that are not observed. We show that the forecasting performance of our model is superior to that of commonly used mixture-amount models, which all assume that the mixture parameters are parametric functions of the amount. The relative performance of benchmark models deteriorates substantially when more amounts are omitted from

the data, but this is much less so for our model. Our model performs best even when half of the observed amount values are omitted, which proves its attractiveness not only for learning the dependence of the mixture parameters on the amount but also for forecasting their values at new amounts.

For this comparison, we consider five parametric models. All these models assume the second-order Scheffé model for the mixture ingredients, which was introduced in Equation (3.2), that is,

$$z_i = \beta_1(A)x_{1i} + \beta_2(A)x_{2i} + \beta_{12}(A)x_{1i}x_{2i} + \sum_{j=1}^{2} \beta_{\text{Region}_\text{j}} d_{\text{Region}_\text{j},i} + \sum_{j=1}^{5} \beta_{\text{Brand}_\text{j}} d_{\text{Brand}_\text{j},i} + \sum_{j=1}^{4} \beta_{\text{Wave}_\text{j}} d_{\text{Wave}_\text{j},i} + \varepsilon_i, \quad (3.29)$$

where $d_{\text{Region}_\text{j},i}$, $d_{\text{Brand}_\text{j},i}$ and $d_{\text{Wave}_\text{j},i}$ are dummy variables which equal one if observation i comes from, respectively, region, brand and wave j, and $\beta_{\text{Region}_\text{j}}$, $\beta_{\text{Brand}_\text{j}}$ and $\beta_{\text{Wave}_\text{j}}$ are the corresponding parameters. The benchmark models differ in the specification of the dependence of the mixture parameters on A. We consider *linear*, *quadratic* and *cubic* functions. The three specifications are

$$\beta_m(A) = \beta_m^0 + \beta_m^1 A, \quad m \in \{1, 2, 12\}, \quad (3.30)$$

$$\beta_m(A) = \beta_m^0 + \beta_m^1 A + \beta_m^2 A^2, \quad m \in \{1, 2, 12\}, \quad (3.31)$$

and

$$\beta_m(A) = \beta_m^0 + \beta_m^1 A + \beta_m^2 A^2 + \beta_m^3 A^3, \quad m \in \{1, 2, 12\}. \quad (3.32)$$

The two final models we include in our comparison are

$$z_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i}x_{2i} + \alpha_1 A + \sum_{j=1}^{2} \beta_{\text{Region}_\text{j}} d_{\text{Region}_\text{j},i} + \sum_{j=1}^{5} \beta_{\text{Brand}_\text{j}} d_{\text{Brand}_\text{j},i} + \sum_{j=1}^{4} \beta_{\text{Wave}_\text{j}} d_{\text{Wave}_\text{j},i} + \varepsilon_i \quad (3.33)$$

and

$$z_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + \alpha_1 A + \alpha_2 A^2 + \sum_{j=1}^{2} \beta_{\text{Region}_\text{j}} d_{\text{Region}_\text{j},i} + \sum_{j=1}^{5} \beta_{\text{Brand}_\text{j}} d_{\text{Brand}_\text{j},i} + \sum_{j=1}^{4} \beta_{\text{Wave}_\text{j}} d_{\text{Wave}_\text{j},i} + \varepsilon_i, \quad (3.34)$$

where we assume that the amount does not affect the mixture parameters but only causes a constant change in the response (see Equation (3.6)). Note that the models in Equations (3.30)-(3.34) have in total 17, 20, 23, 15 and 16 parameters, respectively, that need to be estimated, while our model involves estimating $\mathbf{b}$, τ and $\mathbf{\Phi}$, hence, 10 parameters, to estimate the distribution of the mixture parameters and 11 non-mixture parameters. Therefore, in our model, the total number of parameters to be estimated depends only on the number of mixture ingredients and does not vary with the smoothness of the mixture parameters with respect to the amount.

Our aim is to compare the predictive performance of our model to that of the benchmark models given in Equations (3.30)-(3.34), for campaign recognition at amount values that are omitted during the estimation. To this end, we split the sample into an estimation and a test sample. We compute forecasts for our model by first calculating the posterior predictive distribution, as in Equation (3.7), for the mixture parameters at amount values in the test sample and then combining with the posterior distribution for the non-mixture parameters (see the model in Equation (3.28)). We also estimate the models in Equations (3.30)-(3.34) using Bayesian methodology and use the posterior distributions ($10,000$ draws) of their parameters for forecasting at amount values in the test sample. As a forecasting performance measure, we use the aggregate mean squared error (MSE), which is defined as

$$\text{MSE} = \frac{1}{n} \sum_{j=1}^{n} (\hat{p}_j - p_j)^2,$$

where $\hat{p}_j - p_j$ is the difference between the predicted proportion of recognition for campaign j ($\hat{p}_j$) and the actual proportion (p_j), averaged across all n campaigns in the test sample.

We investigate the forecasting performance using k-fold cross validation with $k = 52/\{4, 13, 26\} = \{13, 4, 2\}$ over the amount values, where we take 4, 13 or 26 *consecutive* amount values for each test sample. We do so in order to withhold parts of the amounts from the

estimation, which makes it more difficult for the models to estimate the pattern of the mixture parameters with respect to the amount. In each round, one of the k subsamples is retained as the test sample and the remaining $k-1$ subsamples are used as the training sample. We calculate the MSE values for all test samples and average them across the k repetitions. We provide the resulting MSE values for our model (GP) and the benchmark models in Table 3.6.

From Table 3.6, we can see that our model performs best for all cases in the k-fold cross validation exercise. For each test set size, it leads to smaller MSEs than the benchmark models. When four amount values are omitted, the MSE value of our model is roughly four times smaller (better) than those of the benchmark models. When we omit 26 amount values (which corresponds to half of the amount values observed in the sample), only one of the benchmark models (model in Equation (3.33)) performs similarly to our model. Note that omitting half of the observed amount values is extreme and is considered here to only evaluate how the models' performances deteriorate if more amount values are held out from the estimation sample. In this case, the forecasting performance of our model stays best. Of the benchmark models, the models with fewer parameters tend to perform better (but not better than our model).

In practice, one could use the Gaussian process prior model to estimate the mixture parameters with respect to the amount variable to obtain some intuition about the possible parametric form. If the mixture parameters resemble a known function in the amount, one may impose it and estimate a standard model like we demonstrated above. This will most likely lead to an improved fit. Then, there is no need for guessing parametric forms for the mixture parameters with respect to the amount and/or formally testing which model fits best.

test set size	GP	linear (Equation (3.30))	quadratic (Equation (3.31))	cubic (Equation (3.32))	A only (Equation (3.33))	A and A^2 (Equation (3.34))
4	0.0107	0.0430	0.0419	0.0437	0.0421	0.0421
13	0.0141	0.0480	0.0508	0.0588	0.0298	0.0422
26	0.0281	0.0417	0.1430	0.1451	0.0282	0.0312

Table 3.6: MSE values for our model based on Gaussian process (GP) and five benchmark models' forecasts in k-fold cross validation when omitting $4, 13$ or 26 consecutive amounts

3.6 Conclusion

In this chapter, we introduced a new flexible but parsimonious model for mixture-amount data. The current approach to model this kind of data involves strong parametric assumptions for the functional form relating the mixture parameters to the total amount variable. Furthermore, when a flexible parameterization is used in the traditional approach, there are many parameters to estimate. The model that we developed does not require any parametric assumptions concerning the relation between the mixture parameters and the amount. Moreover, there is only one parameter that describes how the mixture parameters vary with respect to the total amount.

Our model is based on so-called Gaussian processes and avoids the necessity to a priori specify the shape of the dependence of the mixture parameters on the amount. The Gaussian process is used as a prior on the amount-specific mixture parameters. This prior specifies the correlation between the mixture parameters at different amount values. The strength of this correlation controls the variation of the mixture parameters across different amounts.

We demonstrate that our model outperforms standard models from the literature. As we argue, a parametric function relating the mixture parameters to the amount variable is never known a priori. As a result, we can never be certain whether that parametric assumption is correct. Therefore, in the traditional approach, model comparison and testing procedures are required to choose the final model. This is not needed for the model proposed here. Our modeling approach turns out to be useful to obtain insights in the response to mixture ingredients as well as amounts and has a very good predictive performance.

Chapter 4

Choice Modeling Made More Personal. Extending the ICLV Model With a Gaussian Process Prior

4.1 Introduction

Transportation industry representatives and policy makers are increasingly looking for solutions to negative externalities, such as greenhouse gas emissions and other pollutants, due to mobility of people and goods. Electromobility is considered as a substantial factor in obtaining a sustainable mobility mix (European commission, 2016). Evaluating the impact and potential of such transport in the long-run requires us to analyze and understand the drivers of individual and collective choices for such a technology.

Focusing on the personal car market, electric vehicles (EV) have been on the market for many years. Despite strong advertising and financial incentives, data on observed choices show that, up to few country-specific exceptions, current sales of pure EVs remain somewhat limited. Many reasons can be mentioned, from technology (e.g. vehicle range, limited charging infrastructure) to user habits (e.g. lack of consumer knowledge and experience), to name just a few. To understand which reasons play the largest role and to align industrial strategies with policies in order to achieve sustainability targets, a number of researchers have developed quantitative models that aim at explaining why individuals still prefer internal combustion engine (ICE) cars over EV alternatives.

Academic research on the mathematical modeling of individual demand for electric personal cars dates back to Train (1980); Beggs and Cardell (1980); Beggs et al. (1981), as a response to the oil crisis of the 1970s. The zero-emission vehicle mandate in California also motivated a series of studies on potential demand for EVs, e.g. Bunch et al. (1993); Brownstone et al. (1996); Brownstone and Train (1999); Brownstone et al. (2000). Since then, a lot of studies have been done, see Tanaka et al. (2014) for an inventory of these.

All recent studies have come with advanced mathematical or statistical methods and advanced data collection protocols to give detailed insights in consumer behavior and predicting the impact of the new alternatives on the automotive market structure. Studies that contain empirical estimation of models mostly rely on a combination of stated preference surveys (Hensher et al., 2005) and random utility maximization (RUM) discrete choice models (Ben-Akiva and Lerman, 1985; Train, 2009) or their generalizations (Potoglou and Kanaroglou, 2007; Ben-Akiva et al., 2002; Walker and Ben-Akiva, 2002). Only very few applications, e.g. Brownstone et al. (2000), use revealed preference data.

The attributes which are used in EV choice models are most of the time the same: range, accessibility to charging stations, purchasing price and operating costs. Only more recently, authors started to consider attitudes and opinions as relevant dimensions when understanding choice behavior, see e.g. Walker (2001) for a general framework of analysis and Hurtubia et al. (2014); Temme et al. (2008) or Johansson et al. (2005) for an application to transport mode choice. Such approaches that pertain to modeling demand for electric cars can be found in Beck et al. (2016), who introduced best worst scaling to model attitudes within the choice process; Glerum et al. (2013), who used a hybrid choice model that accounted for attitudes toward leasing contracts and convenience aspects of a vehicle; Daziano and Bolduc (2013), who applied a hybrid choice model to explain environmental preferences in a private vehicle choice context while accounting for attitudinal factors; and Glerum and Bierlaire (2012), who extended a hybrid choice model by capturing the dispersion effects that occur in the answers to attitudinal survey questions. In her review of the literature, Hjorthol (2013) finds that attitudes to and perceptions of electric cars, both positive and negative, vary with experience, knowledge and everyday context. Further, Johansson et al. (2005) point out that a real life complication in traditional modal choice models is individual heterogeneity, beyond what is captured by observed individual specific characteristics. Despite the fact that there are many authors who demonstrated that attitudes and opinions play a role in the choice

process, no clear and robust analysis of how the impact of such attitudes and opinions differ across individuals exists. With this chapter, we attempt to fill this void in the literature.

In this chapter, we develop a choice model that accounts for (latent) environmental consciousness, which may have a heterogeneous impact across the population of decision makers. With respect to new vehicle purchases, environmental consciousness is expected to play a significant role in the individual's decision of whether to buy an EV. As environmental consciousness is not directly observed, we use data on observed indicators acquired through a questionnaire, to infer it for the individuals in the sample. These indicators are answers to a set of attitudinal questions. We assume that the impact of environmental consciousness on the choice of vehicle type depends on a certain characteristic of individuals. In the literature, many studies consistently demonstrate that environmental consciousness is correlated with individuals' age, education and political ideology, see Sánchez and Lafuente (2010) for a summary of these studies. We reason that this shall in turn imply that the impact of environmental consciousness on individual decisions should also depend on individual characteristics. As a result, in our model, we associate the impact of environmental consciousness on the choice of vehicle type to the individual's age, so that policy makers would have some insight into how to personalize incentives for a travel behavior change. However, the developed methodology allows for any other characteristic as well. We also consider technical and price attributes of vehicles, yet, we put special emphasis on understanding if and how environmental consciousness drives choices regarding EVs.

Since not much research has yet looked at associations between individual's age and the impact of his or her environmental consciousness on behavior, we cannot easily specify a function relating the impact of environmental consciousness to individual's age, as we cannot a-priori assume such a parametric form. Therefore, we employ a non-parametric technique, the so-called Gaussian process prior, to infer this dependence from data. Full Bayesian inference is carried out to estimate the actual dependence. We apply our approach to revealed preference data on new vehicles' purchases in France in years 2010 – 2014 and demonstrate the added value of our non-parametric analysis when compared to a model where the parameter of environmental consciousness is assumed to be constant across individuals.

In sum, this chapter is different from the papers mentioned above in two aspects, (1) we present a new choice model that allows heterogeneous effects on the choice across the population with respect to the variable of interest, (2) we use a rich revealed preference data

set, therefore, the results that we present are important in a real green versus conventional vehicle choice situation. The rest of the chapter is organized as follows. Section 4.2 presents the data that we use. Section 4.3 details our modeling approach. Results are discussed in Section 4.4. Finally, we draw our conclusions in the last section.

4.2 Data

The revealed preference data set which we use in this chapter was obtained from a large automobile company. These data were drawn from a pool of five cross-sectional disaggregate surveys regarding new car sales that were carried out from 2010 to 2014 in France. The protocol was the following. Every mid-year, a sample[1] of purchasers of new cars in the former 12 months were sent pre-filled questionnaires[2]. What was pre-filled concerned makes, models and engines of purchased cars. These purchasers were asked to additionally answer a large number of different questions, among which, the price paid for the vehicle, what the car was used for (e.g. work, holidays, bringing children to school, etc), the estimated share of where the car was planned to be used (in town, out of town, motorway), what features it contained (e.g. park assistance system, lane assistance system, USB socket, etc) and socioeconomic and demographic information. In our analysis, we only consider individuals who indicated to have considered buying or who in fact bought an electric or a hybrid vehicle. As a result, the conclusions drawn in this chapter should be interpreted with regard to such a particular sample of individuals. Throughout the chapter, we use the names electric or hybrid vehicle, environmentally friendly vehicle and green vehicle as synonyms.

Since we work with revealed preference data, for the individuals who purchased conventional vehicles, we do not observe what features electric or hybrid cars, that these individuals might have considered, had. In other words, we do not observe attributes for the non-chosen alternatives. Therefore, for the analysis that follows, we shall construct them. For all the vehicle features, except for the price, we use the same values as observed for the corresponding features of the purchased conventional vehicle. In this way, we assume that the individual would have chosen the same options for an electric or hybrid vehicle as he or she chose for a non-electric or non-hybrid vehicle. Put differently, we assume that individuals first decide

[1] Unfortunately, some niche makes of electric cars are not represented in the sample.
[2] For example, in our analysis, 2010 spans from mid 2009 to mid 2010.

on vehicle characteristics other than the engine type, and only then choose between environmentally friendly or conventional vehicles. We have to be more careful with the price variable. We construct a representative price for the electric or hybrid vehicle for the individuals that purchased conventional vehicles by taking an average over the prices of electric and hybrid vehicles that have other features the same as the purchased conventional vehicle.

In our analysis, the dependent variable is binary and is 1 if the individual bought an electric or a hybrid vehicle and 0 otherwise. In Table 4.1, we list the variables that we use in our model as controls with an explanation for each of them. Most of the variables are dummies, with an exception of price, ratios of annual mileage expected to be driven in town/city, out of town and on a motorway, and individual's age. As vehicle characteristics, we consider its price (or representative price, if a purchase of conventional vehicle was observed), whether it has park assistance system, built-in navigation system, sun roof (or full glass roof, panoramic glass roof), lane assistance system, blind spot assistance system and four-wheel drive, and its market category, where we define ten market categories to distinguish among vehicle classes. In our dummy coding, market category 6 (Luxury/Sports) is taken as the benchmark market category. As individual characteristics, we consider one's gender, age, whether one comes from a household of high income class and one's occupation, where we define 11 dummy coded occupation variables. As the benchmark occupation, we consider 'occupation: others'. In Figure 4.1, we provide the distribution of respondents' ages in our sample. The average of all individuals' age in our sample is 49 years. Note that, in our sample, there are relatively fewer observations below 30 or above 70 years of age. Finally, we have information of whether individual uses the new car often to commute to and from work (namely, almost every day) and how much of the annual mileage s/he estimates to be driving in town/city, out of town and on a motorway.

In Table 4.2, for each variable presented in Table 4.1, we provide the average observed values among both environmentally friendly vehicles and conventional vehicles in our sample. As short-hand notations for these averages, we use $\bar{X}_i|\text{green}$ and $\bar{X}_i|\text{conventional}$, respectively, where X_i corresponds to a particular variable. From Table 4.2, we see that most of individuals in our sample use their vehicles to commute to and from work almost every day. However, individuals behave so more when their car is conventional. Next, there are slightly more individuals coming from a high income class among those that purchased a green vehicle. It also appears that the average age is somewhat larger among the indi-

viduals that purchased environmentally friendly vehicles. From the average values of the original price, in our sample, the environmentally friendly vehicles are more expensive than the conventional ones. For more details about the variables, see Tables 4.1-4.2.

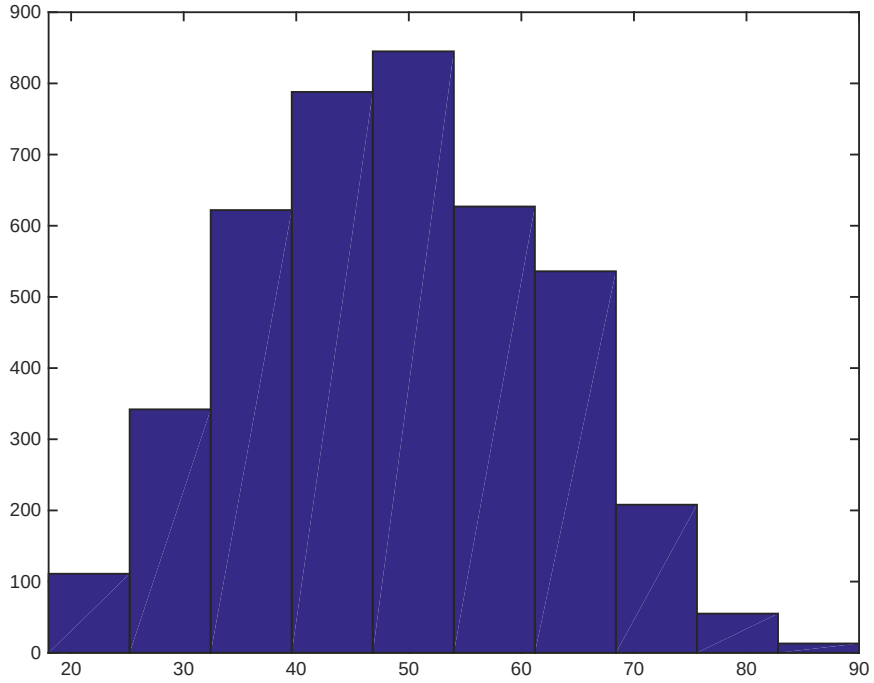


Figure 4.1: Distribution of respondents' age values

Many authors (Hess et al., 2013; Beck et al., 2013; Archtnicht et al., 2011) stress the importance of environmental consciousness in explaining transport-related decisions. As environmental consciousness itself is not directly observed, we use attitudinal variables to infer it for the individuals in the sample. Our data set contains a number of such attitudinal variables, which we refer to as *indicators*. First, in the questionnaire, respondents had to evaluate several statements on a four-point Likert scale, from strongly agree (4) to strongly disagree (1). These statements are intended to measure respondents' environmental consciousness. The statements are as follows:

- To help the environment, I bought a car with lower CO2 emissions (I_1)
- I chose a smaller car, for environmental reasons (I_2)
- I chose a smaller engine, for environmental reasons (I_3)

X_i	Description
commute to and from work	The new car is used to commute to and from work *almost every day*
occupation: company owner ($\geq$ 10 emp.)	Owner of a company with 10 or more employees
occupation: company owner ($<$ 10 emp.)	Owner of a small business, shop, farm etc with less than 10 employees
occupation: professional (self employed)	e.g. Doctor, Lawyer, Architect, Accountant etc
occupation: professional (employed)	Professional/Senior Civil Servant/Company Director/ Senior Manager
occupation: tradesman/craftsman	e.g. Builder, Plumber, Painter, Electrician etc
occupation: manager/supervisory	Manager/Supervisory
occupation: public/private services	Healthcare, Teaching, Police, Armed Forces etc
occupation: clerical	Clerical Worker
occupation: manual	Manual Worker
occupation: others	Housewife/Househusband, Student, Retired, Without Work/Unemployed/None of the Above
high income class	household annual gross income is $> 50,000$ Euros
male	if male
age	individual's age (18 - 90)
price	The total price paid in thousands of Euros
market category 1	Econbox
market category 2	Small
market category 3	Lower medium
market category 4	Medium
market category 5	Full
market category 6	Luxury/Sports
market category 7	Off Road
market category 8	Small MPV (multi-purpose vehicle)
market category 9	Large MPV (multi-purpose vehicle)
market category 10	LUV (light utility vehicle)
park assistance	vehicle has park assistance system
navigation	vehicle has built-in navigation system
sun roof	vehicle has sun roof/full glass roof/panoramic glass roof
lane assistance	vehicle has lane assistance system
blind spot assistance	vehicle has blind spot assistance system
4 wheel drive	four-wheeled vehicle
	What is the estimated ratio of your annual mileage for each type of driving listed below:
ratio of driving in town/city	Driving in town/city
ratio of out of town driving	Out of town driving
ratio of motorway driving	Motorway driving

Table 4.1: Explanatory variables with their descriptions on the right hand side

X_i	$\bar{X}_i$|green	$\bar{X}_i$|conventional
commute to and from work	0.56	0.69
occupation: company owner ($\geq$ 10 emp.)	0.07	0.03
occupation: company owner ($<$ 10 emp.)	0.07	0.04
occupation: professional (self employed)	0.12	0.09
occupation: professional (employed)	0.16	0.09
occupation: tradesman/craftsman	0.03	0.02
occupation: manager/supervisory	0.22	0.25
occupation: public/private services	0.18	0.22
occupation: clerical	0.10	0.18
occupation: manual	0.03	0.05
occupation: others	0.02	0.03
high income class	0.56	0.46
male	0.81	0.73
age	53	47
original price	30.89	22.94
representative price	30.89	30.20
market category 1	0.0035	0.0450
market category 2	0.18	0.36
market category 3	0.49	0.21
market category 4	0.12	0.09
market category 5	0.02	0.02
market category 6	0.05	0.01
market category 7	0.08	0.12
market category 8	0.05	0.10
market category 9	0.0009	0.0194
market category 10	0.0052	0.0234
park assistance	0.62	0.50
navigation	0.63	0.47
sun roof	0.37	0.27
lane assistance	0.06	0.08
blind spot assistance	0.04	0.08
4 wheel drive	0.23	0.11
ratio of driving in town/city	0.35	0.34
ratio of out of town driving	0.36	0.40
ratio of motorway driving	0.30	0.27

Table 4.2: Explanatory variables with their average values observed among green vehicles (second column) and conventional vehicles (third column) in our sample

- For me cars are causing many of the environmental problems (I_4)
- I would pay more for environmentally friendly features (I_5)

Further, in the questionnaire, respondents had to write down answers about their newly bought car to the following open questions:

- Please list here all the good points of your new car, the things you particularly like about it
- If there are any points on which you are dissatisfied, please detail them
- For what reasons did you buy this particular new car rather than some other one?
- What were the reasons why you decided not to buy this *main alternative* car?
- Why didn't you choose the same make as before?

We hypothesize that if someone (himself or herself) indicated an environmentally friendly feature for the above questions, he or she is (to a certain extent) environmentally conscious. In the questionnaire, there were one to two environmentally friendly features for each question above named by respondents: polluting/non-polluting and recyclable/non-recyclable. We consider these questions as binary indicators (if environmentally friendly feature named (1) or not (0) by respondent) and include them in the model in the same manner as the indicators $I_1 - I_5$. We provide the resulting binary indicators together with the indicators $I_1 - I_5$ in Table 4.3 with the corresponding average values observed in our sample among green vehicles and among conventional vehicles, in the same manner as for the variables in Table 4.2. As a result, in the final model, there are 11 indicators in total, $I_1 - I_{11}$, five four-point Likert scale indicators and six 0/1 indicators.

I_k	$\bar{I}_k$|green	$\bar{I}_k$|conventional
lower CO2 (I_1)	1.40	2.04
smaller car (I_2)	2.53	2.68
smaller engine (I_3)	2.46	2.57
cause env. problems (I_4)	2.29	2.39
would pay more (I_5)	1.97	2.20
like: low pollution (I_6)	0.0539	0.0063
like: recyclable (I_7)	0.0009	0
dislike: polluting (I_8)	0	0.0020
reason: low pollution (I_9)	0.15	0.03
reason reject: polluting (I_{10})	0.0165	0.0043
reason change: polluting (I_{11})	0.0270	0.0047

Table 4.3: Indicators with their average values observed among green vehicles (second column) and conventional vehicles (third column) in our sample

An interesting point to note from Table 4.3 is that, in our sample, for the indicators where one had to only tick the corresponding box that measured his or her agreement with

the statement ($I_1 - I_5$), the respondents who did not purchase an environmentally friendly vehicle scored higher values (that is, indicated a stronger agreement) than the respondents who bought an environmentally friendly vehicle. However, in the open questions ($I_6 - I_{11}$), the respondents who purchased a green vehicle scored better (that is, named environmentally friendly features more often) than the respondents who purchased a conventional vehicle. This may seem strange, however, it is in line with the literature (see, for example, Grimm (2010)) where many authors point out that when evaluating attitudinal statements relating to the environment, people feel some (inner) pressure to demonstrate high(er) environmental friendliness, while in reality they do not necessarily behave accordingly. It seems that we observe such a scenario in our sample, where the individuals who did not buy a green vehicle feel some sort of pressure (or guilt) and try to demonstrate how environmentally conscious they are by indicating a stronger agreement to the attitudinal statements. However, since open questions are less obvious in what they intend to measure, these individuals fail to name environmentally friendly features in their answers to the open questions.

After omitting the missing values for the variables described above, the final sample contains $4,147$ observations ($\approx 29\%$ of the initial data set), with $1,150$ observed purchases of electric or hybrid vehicles, which correspond to 149, 181, 137, 302 and 381 purchases in years 2010 – 2014, respectively. In terms of percentages, this constitutes 26.51%, 29.53%, 32.78%, 49.27% and 19.63% of all purchases in years 2010 – 2014, respectively. In the analysis that follows, we will include year fixed effects to capture the differences in purchasing rates in our sample.

4.3 Methodology

4.3.1 Model

To model the choice between environmentally friendly and conventional vehicles, we build on the Bayesian analysis of the binary probit model as in Albert and Chib (1993). It is based on the concept of latent utility maximization, where the alternative with the highest utility is chosen (Ben-Akiva and Lerman, 1985; Train, 2009). Let the latent utility of an

environmentally friendly vehicle for individual n, $n = 1, 2, \ldots, N$, be

$$y_n^* = X_n'\beta + \gamma(age_n)ENV_n + \varepsilon_n, \quad \varepsilon_n \sim N(0,1), \tag{4.1}$$

where X_n contains explanatory variables and β is the corresponding parameter vector. We assume that the individual's utility also depends on his or her environmental consciousness, ENV_n, which is not directly observed. Its impact on the utility is given by $\gamma(age_n)$ and depends on individual's age. If y_n^* is positive, individual n will choose an environmentally friendly vehicle, i.e.,

$$y_n = \begin{cases} 1, & \text{if } y_n^* > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{4.2}$$

where y_n is a binary variable equal one in case an environmentally friendly vehicle is chosen and zero otherwise. Instead of observing individual's environmental consciousness, we only observe his or her responses to attitudinal questions about the environment. These attitudinal questions ask to indicate to which extent the respondent agrees with a particular statement that relates to the environment. The obtained indicators can be expected to be important when explaining (green) vehicle ownership, however, including them directly in the choice model would lead to the risk of measurement error and, therefore, endogeneity bias. The reason is that answers to attitudinal questions are not direct measures of attitudes but rather functions of underlying latent attitudes. In order to jointly model the observed choices and responses to attitudinal questions, we extend the simple binary probit model to the Integrated Choice and Latent Variable (ICLV) model as in Hess and Beharry-Borg (2012), for example. In the first step, we relate the indicators of the attitudinal questions to the (latent) environmental consciousness by writing

$$I_{nk} = \delta_{I_k} + \lambda_{I_k} ENV_n + \upsilon_{nk}, \quad \upsilon_{nk} \sim N(0, \sigma_{I_k}^2), \tag{4.3}$$

where I_{nk} is the value for the k^{th}, $k = 1, 2, \ldots, K$, attitudinal indicator for individual n, δ_{I_k} is a constant for the k^{th} indicator, λ_{I_k} is the estimated effect of the underlying environmental consciousness ENV_n on this indicator and υ_{nk} is an error term which is assumed to be independent and normally distributed with zero mean and variance $\sigma_{I_k}^2$. Denote $\Theta = \text{diag}(\sigma_{I_1}^2, \sigma_{I_2}^2, \ldots, \sigma_{I_K}^2)$. We follow common practice in the literature and use a con-

tinuous assumption for the discrete indicators in Equation (4.3) (see, for example, Daziano (2010); Hess and Beharry-Borg (2012)). Further, as noted by Daziano (2010), indicators in Equation (4.3) are highly correlated. This correlation should be completely captured by the environmental consciousness ENV_n, yielding the covariance between two indicators I_{nk} and I_{nl} given by $\lambda_{I_k}\lambda_{I_l}$. In addition, the assumption of independent error terms in Equation (4.3) is required for identification of the model, as otherwise the correlation between the indicators would not be captured by ENV_n but rather taken up by υ_{nk}'s.

The latent environmental consciousness is also linked to observed sociodemographic characteristics as follows

$$ENV_n = z_n'\alpha + \zeta_n, \quad \zeta_n \sim N(0,1), \tag{4.4}$$

where z_n contains sociodemographic variables for individual n with the corresponding parameter vector α and ζ_n is an error term which is assumed to follow a standard normal distribution. The variance of one for the errors is required for identification of λ_{I_k}, $\forall k$, in Equation (4.3). Note that X_n in Equation (4.1) may contain variables that are contained in z_n. In the structural equation modeling literature, Equations (4.1) and (4.4) are called structural equations and Equations (4.2) and (4.3) are called measurement equations.

Let us summarize our model as follows:

Measurement equations

$$y_n = \begin{cases} 1, & \text{if } y_n^* > 0, \\ 0, & \text{otherwise,} \end{cases}$$

$$I_{nk} = \delta_{I_k} + \lambda_{I_k} ENV_n + \upsilon_{nk}, \quad \upsilon_{nk} \sim N(0, \sigma_{I_k}^2), \quad k = 1, 2, \ldots K.$$

Structural equations

$$y_n^* = X_n\beta + \gamma(age_n)ENV_n + \varepsilon_n, \quad \varepsilon_n \sim N(0,1),$$
$$ENV_n = z_n'\alpha + \zeta_n, \quad \zeta_n \sim N(0,1).$$

We also illustrate the structure of the ICLV model in Figure 4.2, where observed components are given in rectangles and unobserved components are given in ellipses. Respondent char-

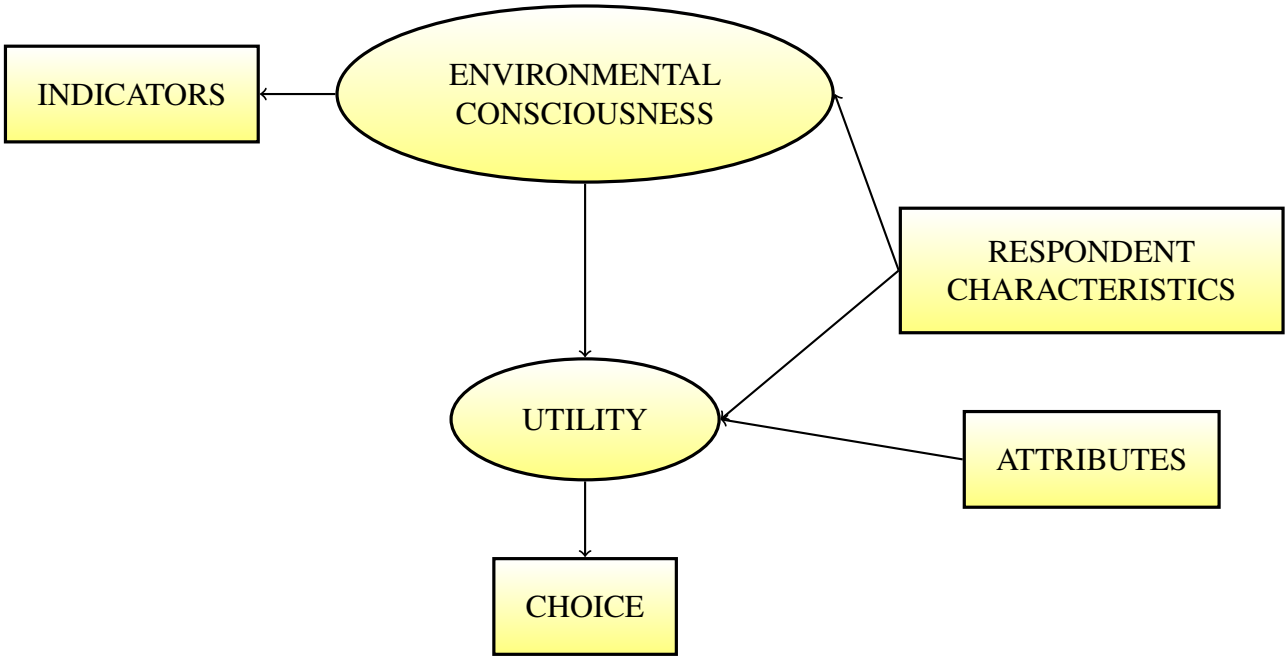


Figure 4.2: Structure of ICLV model

acteristics affect both the latent environmental consciousness and the latent utility, where the latter also depends on measured vehicle attributes. The latent environmental consciousness affects the response to the attitudinal questions (providing indicators) and also the latent utility which in turn affects the probability of choosing an environmentally friendly vehicle.

Finally, we assume that the effect of the latent environmental consciousness on the latent utility in Equation (4.1) is some function of individual's age, that is, we write $\gamma(age)$, where the effect for individual n is $\gamma(age_n)$. For this function, we only assume that the values of γ are similar for individuals with a similar age and not necessarily so for individuals with a (very) different age. How similar or dissimilar parameters γ are with respect to individuals' age, will be determined from the data by estimating a single parameter τ. To complete the formal model specification, we assume that the prior on $\gamma(age)$ is a Gaussian process (Bishop, 2006; Rasmussen and Williams, 2005) with mean g and variance-covariance matrix $\mathbf{\Sigma}$ or $\gamma(\overrightarrow{age}) \sim N(g, \mathbf{\Sigma})$, where we denote $\overrightarrow{age} = (age_1, age_2, \ldots, age_r)'$, $r \leq N$, the set of unique observed age values in the data. Further, $g = g^* \iota_{r\times 1}$ with a scalar g^* and ι denotes a vector of ones. In other words, on average, the parameter of environmental consciousness equals g^*, however, at every age, the parameter deviates from this average. The variance-covariance matrix $\mathbf{\Sigma}$ is further decomposed into a scale factor and a correlation matrix, that

is, $\boldsymbol{\Sigma} = \phi\boldsymbol{\Omega}$. Here, $\boldsymbol{\Omega}$ is a correlation matrix, which specifies the correlation structure for the environmental consciousness parameter γ across different values of age, that is,

$$\mathrm{Corr}\big(\gamma(\overrightarrow{age})\big) = \boldsymbol{\Omega},$$

where a representative element of $\boldsymbol{\Omega}$ is specified as

$$\boldsymbol{\Omega}(age', age'') = \exp(-\frac{1}{2\tau^2}||age' - age''||^2), \quad \tau > 0. \tag{4.5}$$

The expression above is a squared exponential (Gaussian) kernel evaluated at two age values, age' and age''. For any pair of age values, age' and age'', for given τ, a Gaussian process with this correlation function implies:

- $\gamma(age')$ and $\gamma(age'')$ will tend to have high correlation if age' and age'' are "close" to each another, since $||age' - age''|| \approx 0 \implies \boldsymbol{\Omega}(age', age'') = \exp(-\frac{1}{2\tau^2}||age' - age''||^2) \approx 1$,
- $\gamma(age')$ and $\gamma(age'')$ will tend to have low correlation if age' and age'' are "far" apart, since $||age' - age''|| \gg 0 \implies \boldsymbol{\Omega}(age', age'') = \exp(-\frac{1}{2\tau^2}||age' - age''||^2) \approx 0$.

In other words, functions drawn from a Gaussian process with the Gaussian kernel will be locally smooth with high probability. Therefore, the parameter γ values for the similar age values will also be similar and this similarity will decrease as a function of distance between age' and age''. The parameter τ determines how smoothly $\gamma(age)$ varies with age, where different values of τ produce different scenarios. After estimating τ, the inferred shape of $\gamma(age)$ may provide valuable insights into the impact of environmental attitudes of individuals as a function of their age (or another characteristic of interest), which is important in order to align policies and personalize the incentives for various target groups for a travel behaviour change.

As two extreme cases, values of τ equal to zero or to a large positive value result in more standard models: if $\tau \to 0$, we actually allow for separate γ coefficients for each unique value of age, as $\tau \to 0 \implies \boldsymbol{\Omega}(age', age'') = \exp(-\frac{1}{2\tau^2}||age' - age''||^2) \to 0$ for $age' \neq age''$. If we let $\tau \to \infty$, we impose a constant γ coefficient (i.e., independent of age), as $\tau \to \infty \implies \boldsymbol{\Omega}(age', age'') = \exp(-\frac{1}{2\tau^2}||age' - age''||^2) \to 1$. By choosing τ values

between 0 and ∞, we can describe settings in between constant and separate γ parameters with respect to age, without restricting ourselves to any particular function for γ. For more details, we refer to Chapter 3.

4.3.2 Estimation

For estimating the parameters in our model, we use Bayesian techniques. Bayesian inference requires taking draws from the joint posterior distribution of the model parameters (Bishop, 2006; Gelman et al., 2013; Greenberg, 2014; Zellner, 1996). This means that, in order to estimate the parameters β,$\gamma(\overrightarrow{age})$, g, τ, ϕ, α, δ_{I_k}, λ_{I_k} and $\sigma^2_{I_k}$, $\forall k$, we have to sample from the joint posterior density $p(\beta, \gamma(\overrightarrow{age}), g, \tau, \phi, \alpha, \delta_{I_k}, \lambda_{I_k}, \sigma^2_{I_k}|y_n, I_{nk}, \forall n, k)$. Since the environmental consciousness, ENV_n, is not observed, we sample ENV_n alongside the model parameters. Further, to estimate the binary probit model, we have to sample the latent utilities y^*_n in Equation (4.1) as well and consider $p(ENV_n, y^*_n, \beta, \gamma(\overrightarrow{age}), g, \tau, \phi, \alpha, \delta_{I_k}, \lambda_{I_k}, \sigma^2_{I_k}|y_n, I_{nk})$. Sampling latent constructs (latent environmental consciousness and latent utilities in our case) is known as data augmentation (Tanner and Wong, 1987). We choose for a conservative approach and sample the environmental consciousness conditioned only on the indicators I_{nk} but not on the choices y_n. In this way, we only allow one's environmental consciousness to be summarized by the attitudinal indicators and not by the actual choices of vehicles. Although this is suboptimal in case the model structure is correct, we avoid the possibility of a self-fulfilling prophecy where the environmental consciousness is inferred in such a way that it automatically explains vehicle choice. We examined if conditioning on y_n changes the empirical results. However, we find no substantial difference in the results.

Sampling from the joint posterior density of the model parameters is not directly feasible, therefore, we employ a Gibbs sampler (Casella and George, 1992), where a Metropolis-Hastings (Chib and Greenberg, 1995) step is necessary for τ. We derive the sampling distributions for all the parameters and the latent constructs in Appendix 4.A.

4.4 Results

In this section, we present the estimation results. In the analysis, we retained only those variables, whose parameters' highest posterior density (HPD) intervals did not contain the

value zero. We constructed an HPD interval for each parameter such that it is the smallest interval containing 95% of its posterior mass.

Prior to analysis, we rescale the age variable by dividing it by its standard deviation, in order to make the specification of the prior distribution for τ less dependent on the scale of age. We then consider $p(\tau) \sim \ln N(0.2, 0.1)$ as a prior for τ. Note that specifying a prior for τ is in fact specifying a prior on the correlation structure across different age values. Therefore, we choose $p(\tau)$ such that we keep the prior on the environmental consciousness parameters across different age values away from extreme correlations. Further, as priors for $(g \quad \beta')'$ and ϕ, we consider $N(0, 5\,\mathbf{1})$ and $IG2(1, 5)$, respectively, where $\mathbf{1}$ denotes an identity matrix. Note that five degrees of freedom of the prior distribution for ϕ is the minimum required for the expected value and variance of ϕ to exist. Note further that the model specifies a fixed scale for the latent utility and the environmental consciousness by restricting the variance of the error terms to one, see Equations (4.1) and (4.4). As a result, the prior variance of five for g and β corresponds to a rather uninformative prior. The aforementioned priors imply the prior $\gamma(\overrightarrow{age}) \sim N(g, \Sigma)$ with $\Sigma = \phi\Omega$. Hence, the prior on the environmental consciousness' effect on the latent utility specifies a quite uncertain average value across the population. It further allows for variation with respect to individuals' ages by allowing ϕ to differ from zero. The prior specifies that the environmental consciousness parameters are similar (only) for individuals with a similar age.

Next, observe that ENV_n is a latent construct. For interpretability, we want this measure to be positively related to the true environmental consciousness of individuals. We code this in our prior by setting $\lambda_{I_k} \sim N(1, 1)$. This means that we expect ENV_n to be positively related to the observed indicators. Finally, we consider $\alpha_i \sim N(0, 1)$, $\forall i$, $\delta_{I_k} \sim N(1, 1)$, $\forall k$ and we use diffuse priors for $\sigma^2_{I_k}$, that is, $p(\sigma^2_{I_k}) \propto \sigma^{-2}_{I_k}$, $\forall k$.

We used $10,000$ iterations in a Gibbs sampler and the following results are presented after removing $4,000$ initial draws as a burn-in, where we obtained an acceptance rate of 48.64%. The posterior means, standard deviations and 95% HPD intervals for δ_{I_k} and λ_{I_k}, $\forall k$, α and β are given in Tables 4.4, 4.5 and 4.6, respectively. To demonstrate convergence, we plot the full chain of $10,000$ draws for τ in Figure 4.3. Its posterior mean, standard deviation and 95% HPD interval after a burn-in of $4,000$ draws are 1.2781, 0.1256 and $[1.0292$ $1.5126]$, respectively. We demonstrate the implied correlation structure for $\gamma(\overrightarrow{age})$ across different age values at the posterior mean of τ by plotting this correlation versus the differ-

ences in the original age values (top axis) and rescaled age values (bottom axis) in Figure 4.4. Finally, we plot the posterior mean of $\gamma(\overrightarrow{age})$ versus both the original age (top axis) and the rescaled age (bottom axis) values in Figure 4.5, where we also provide the 95% HPD interval (green lines above and below the blue line in the middle. The blue line corresponds to the posterior mean of $\gamma(\overrightarrow{age})$). The horizontal red line corresponds to the posterior mean of g. The dots on the curves denote the observed age values in our data set.

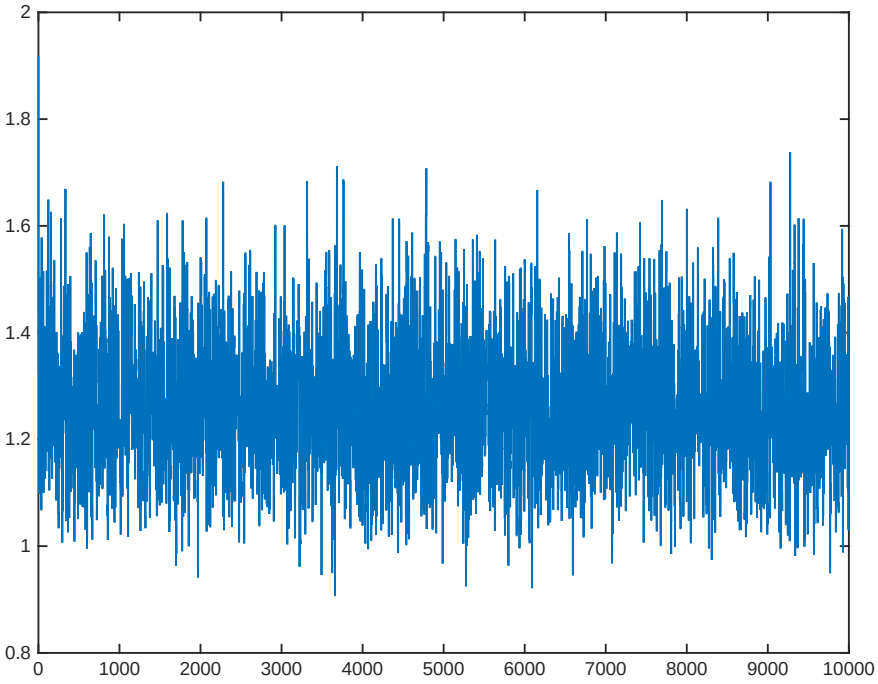


Figure 4.3: Posterior τ draws

Next, we discuss the estimated parameters. We begin with the estimates of the parameters in Equation (4.1) that are given in Table 4.6. From the table, we see that, ceteris paribus, the utility of choosing an environmentally friendly vehicle increases with one's age. Further, respondents who use the vehicle to commute to and from work, ceteris paribus, are less likely to choose an environmentally friendly vehicle, while respondents that have higher-end occupations, all else being equal, are more likely to choose such a vehicle. While controlling for other variables in the model, people tend to choose environmentally friendly vehicles if they drive more in town, whereas neither out of town nor motorway driving have an effect on the utility. This may be partially explained by the fact that there are more charging facilities in town as opposed to out of town or on motorways and, further, distances driven

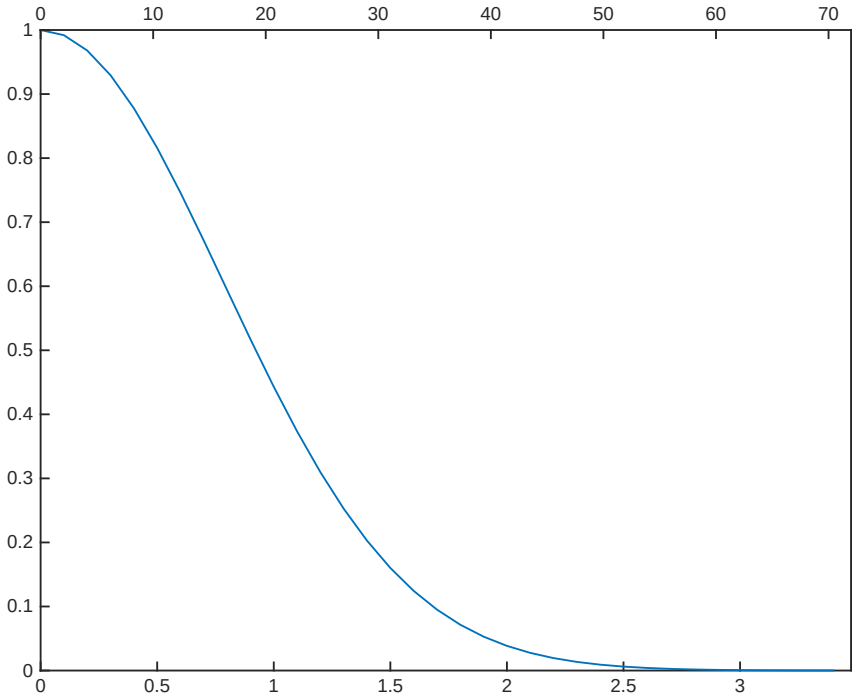


Figure 4.4: Implied correlation at the posterior mean of τ versus the differences in age (top axis) and the differences in rescaled age (bottom axis)

in town tend to be shorter. All else being constant, the price has a negative effect on the latent utility. All market category dummies have strong negative effects meaning that if an individual is interested in any other market category than Luxury/Sports (base category), s/he is less likely to choose a green vehicle, all else being equal. Next, blind spot assistance system has a strong and negative effect on the utility, however, we shall be a bit cautious when interpreting this effect as the proportion of vehicles that have this feature is very low in out data set (see Table 4.1). Further, ceteris paribus, navigation system, sunroof (and/or full glass roof, panoramic glass roof) and four-wheel drive all have positive effects on the utility of choosing environmentally friendly vehicle. Finally, all included year fixed effects capture significant variation in the purchase rate of green vehicles in our sample (see Section 4.2).

In Table 4.4, we can see the indicators to which environmental consciousness is related. Environmental consciousness loads positively on all the indicators meaning that an increase in environmental consciousness results in an increase in the degree to which respondents agree with the statements relating to the environment or the degree to which they match their expectations about a car to environmentally friendly features. The highest values are

obtained for λ_{I_2}, λ_{I_3} (among the four-point Likert scale indicators) and λ_{I_9} (among the 0/1 indicators), which means that the indicators I_2, I_3 and I_9 are most strongly correlated with the latent environmental consciousness. The posterior mean of the variance matrix Θ is $\mathbb{E}(\Theta|y_n, I_{nk}) = \mathrm{diag}(0.5892, 0.2367, 0.1990, 0.6294, 0.5466, 0.0190, 0.0559, 0.0076, 0.0107)$. This shows that not all of the attitudinal indicators are related to the environmental consciousness to the same extent. We can see that, among the four-point Likert scale indicators, I_3 has the lowest error variance and, among the 0/1 indicators, I_{10} has the lowest error variance. Thus, the indicators I_3 and I_{10} are related to the environmental consciousness most.

I_k	$\mathbb{E}(\delta_{I_k}\|y)$ (StDev$(\delta_{I_k}\|y)$)	$\mathbb{E}(\lambda_{I_k}\|y)$ (StDev$(\lambda_{I_k}\|y)$)	95% HPD interval δ_{I_k}		λ_{I_k}	
lower CO2 (I_1)	3.0017 (0.0443)	0.4788 (0.0134)	[2.9258	3.0926]	[0.4528	0.5048]
smaller car (I_2)	2.0848 (0.0851)	0.9568 (0.0138)	[1.9415	2.2645]	[0.9324	0.9861]
smaller engine (I_3)	2.1747 (0.0873)	0.9846 (0.0135)	[2.0318	2.3643]	[0.9582	1.0110]
cause env. problems (I_4)	2.5616 (0.0263)	0.2573 (0.0133)	[2.5145	2.6172]	[0.2304	0.2829]
would pay more (I_5)	2.7976 (0.0241)	0.2361 (0.0122)	[2.7535	2.8461]	[0.2115	0.2592]
like: low pollution (I_6)	0.0156 (0.0025)	0.0135 (0.0022)	[0.0105	0.0204]	[0.0091	0.0179]
reason: low pollution (I_9)	0.0488 (0.0056)	0.0449 (0.0039)	[0.0386	0.0607]	[0.0372	0.0524]
reason reject: polluting (I_{10})	0.0061 (0.0015)	0.0055 (0.0014)	[0.0032	0.0090]	[0.0027	0.0082]
reason change: polluting (I_{11})	0.0086 (0.0018)	0.0078 (0.0017)	[0.0049	0.0119]	[0.0045	0.0109]

Table 4.4: Estimates of the parameters δ_{I_k} and λ_{I_k}, $\forall k$, in Equation (4.3)

Table 4.5 shows the estimates of the parameters in Equation (4.4), explaining how the environmental consciousness depends on individual characteristics. From this table, we conclude that respondents from a high income class and males tend to be less environmentally conscious on average, while one's age has a positive effect on his/her environmental consciousness. All included occupation dummies have positive effects on the environmental consciousness (recall that the baseline is respondents with no income, such as, housewives/househusbands, students, unemployed, etc). To further examine if one's occupation

itself matters in determining how environmentally conscious one is, we calculate the HPD intervals for differences in parameters' posterior draws for all occupation pairs. If such an interval for a particular pair of parameters contains a zero value, this implies that the corresponding occupations do not have significantly different effects on the environmental consciousness. Out of all the occupations in Table 4.5, being a manager or having a supervisory job implies a significantly smaller environmental consciousness than working in public or private services, all else being equal. The remaining occupations did not have significantly different effects on the environmental consciousness.

z_i	$\mathbb{E}(\alpha_i\|y)$ (StDev$(\alpha_i\|y)$)	95% HPD interval	
high income class	-0.1470 (0.0359)	$[-0.2186$	$-0.0775]$
male	-0.2051 (0.0398)	$[-0.2822$	$-0.1255]$
age	0.0072 (0.0013)	[0.0045	0.0097]
occupation: professional (self employed)	0.1623 (0.0720)	[0.0261	0.3077]
occupation: tradesman/craftsman	0.2239 (0.1137)	[0.0120	0.4533]
occupation: professional (employed)	0.1787 (0.0680)	[0.0432	0.3130]
occupation: manager/supervisory	0.1321 (0.0581)	[0.0176	0.2449]
occupation: public/private services	0.2275 (0.0595)	[0.1128	0.3428]
occupation: clerical	0.1966 (0.0652)	[0.0621	0.3167]
occupation: manual	0.2674 (0.0943)	[0.0843	0.4525]

Table 4.5: Estimates of the parameters α in Equation (4.4)

We next turn our attention to the estimated function $\gamma(\overrightarrow{age})$, where, in Figure 4.5, we plot its posterior mean (blue middle curve) together with the 95% HPD interval (green curves outside the inner curve). From the figure, we can see that the environmental consciousness increases the utility of choosing environmentally friendly vehicle by, on average, $\mathbb{E}(g|y_n, I_{nk}) = 0.2640$ (horizontal red line). The effect varies with age, however. For the youngest and oldest individuals in the sample, the HPD intervals are too wide to make con-

clusions (recall that we have relatively few observations for the youngest and oldest individuals in the population, see Figure 4.1). For individuals in their 40s-60s (namely, 42 - 60 years of age), the effect of environmental consciousness on the utility of environmentally friendly vehicle is significantly lower than the average estimated effect for the population.

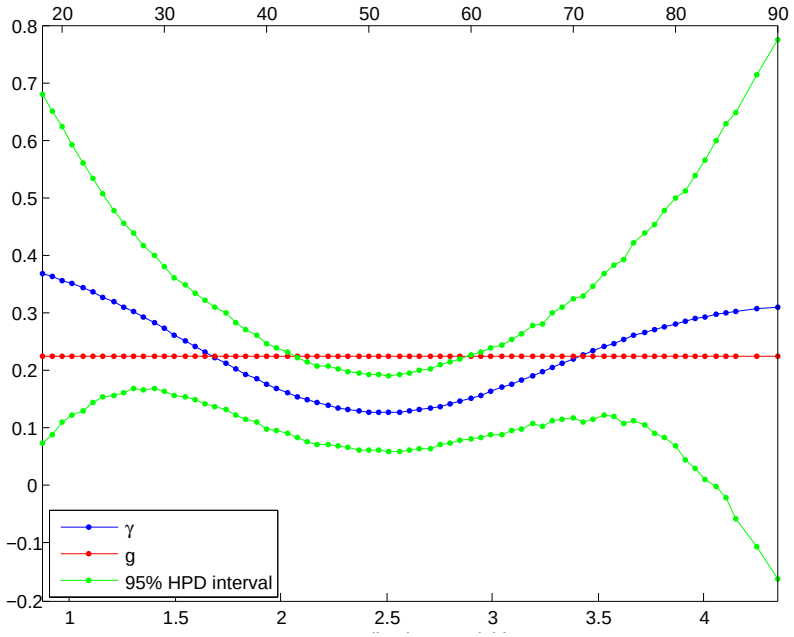


Figure 4.5: Posterior mean of $\gamma(\overrightarrow{age})$ with 95% HPD interval. The horizontal red line is the posterior mean of g

This result signals two potential policies for this age group. First, one could foster these individuals' environmental consciousness (by personalized mailing with an emphasis on the importance of saving the environment, for example). However, it might be that these individuals are actually environmental conscious just that they do not behave accordingly when it comes to choosing a vehicle. If this is the case, we instead need to demonstrate to them how much of the environment they would save by using a green vehicle instead of a conventional one, for example, again by personalized mailing, which leads to a second (and essentially different) policy. To choose between the two policies, one needs to know the environmental consciousness of individuals, which is not observed. However, it becomes observable through a sampling step in our Bayesian framework. Therefore, we next examine how the latent environmental consciousness varies with respect to individual's age. Note first from Table 4.5, that the effect of age on the latent environmental consciousness is 0.0072, which is relatively small. To examine this further, in Figure 4.6, we provide box plots for

the posterior draws of the latent environmental consciousness per age category, where each box corresponds to a separate age group, indicated on the horizontal axis. The diamonds on each box mark the posterior means of the environmental consciousness for each age category. From the posterior means, it is clear that individuals from the categories comprising $42 - 60$ years of age are more or less on a par with other age groups in terms of the level of environmental consciousness, irrespective of their choices with respect to green vehicles, which indeed corresponds to the small effect of age on the latent environmental consciousness obtained above. This further implies that, for this age group, policy advisors should not concentrate on fostering environmental friendliness, as the respective individuals are not lacking it anyway. Policy advisors should rather focus on introducing environmentally friendly vehicles to these individuals and explaining how they contribute in behalf of the environment. This would appeal to individuals' environmental attitudes and should in turn lead to more purchases of environmentally friendly vehicles.

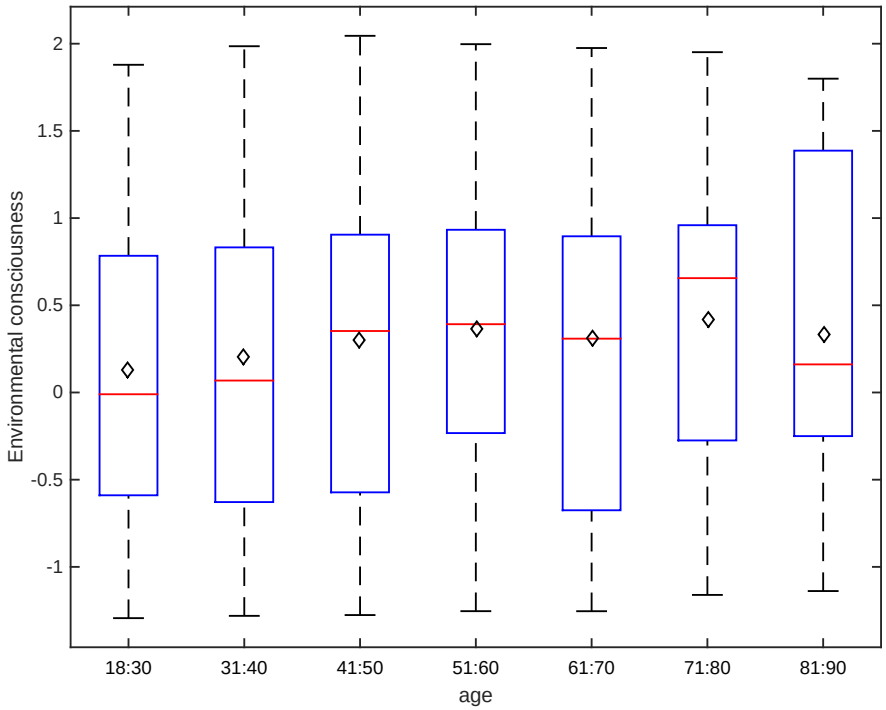


Figure 4.6: Boxplots for the posterior draws of the environmental consciousness versus age categories. The diamonds mark the posterior means for the corresponding age categories

Finally, the posterior mean and standard deviation of ϕ are, respectively, $\mathbb{E}(\phi|y_n, I_{nk}) = 0.5014$ and $\mathrm{StDev}(\phi|y_n, I_{nk}) = 0.3149$, demonstrating that there is a lot of variation among different ages with respect to their environmental consciousness.

The above analysis demonstrates how policy advisors could exploit the information provided by our model. In our application, we chose age as the characteristic to be used to differentiate individuals. However, any other characteristic could be chosen. The analysis can then be performed for each group of individuals in a similar manner as we demonstrated above. Since our data did not allow us to elaborate about the individuals younger than 42 and older than 60, we only provided an example of interpreting the results for the age group of $42 - 60$. However, given more data, this analysis could be repeated for each individual category.

The type of information that we presented above is important when designing personalized incentives for drivers. As people are different, in many cases, different groups of individuals demand different attention. In our application, it would be wise to design incentives targeted to the individuals in their 40s-60s that would be meant to explain why green vehicles are better for the environment. Such personalized policies could not be achieved if the environmental consciousness parameter would be assumed to be equal for all individuals.

X_i	$\mathbb{E}(\beta_i \mid y)$ (StDev$(\beta_i \mid y)$)	95% HPD interval
intercept	-0.2112 (0.2354)	$[-0.6641 \quad 0.2638]$
age	0.0137 (0.0022)	[0.0094 0.0180]
commute to and from work	-0.1790 (0.0590)	$[-0.2994 \; -0.0693]$
occupation: company owner (≥ 10 emp.)	0.3686 (0.1187)	[0.1311 0.5966]
occupation: company owner (< 10 emp.)	0.4220 (0.1115)	[0.1967 0.6324]
occupation: professional (self employed)	0.2262 (0.0823)	[0.0704 0.3927]
occupation: professional (employed)	0.1956 (0.0767)	[0.0423 0.3372]
price	-0.0172 (0.0032)	$[-0.0234 \; -0.0111]$
market category 1 (Econbox)	-3.1211 (0.2767)	$[-3.6562 \; -2.5781]$
market category 2 (Small)	-1.8776 (0.1611)	$[-2.1972 \; -1.5602]$
market category 3 (Lower medium)	-0.9281 (0.1540)	$[-1.2230 \; -0.6183]$
market category 4 (Medium)	-1.5197 (0.1660)	$[-1.8397 \; -1.1844]$
market category 5 (Full)	-1.4314 (0.2137)	$[-1.8502 \; -1.0064]$
market category 7 (Off Road)	-2.3174 (0.1730)	$[-2.6465 \; -1.9707]$
market category 8 (Small MPV)	-1.7916 (0.1732)	$[-2.1213 \; -1.4469]$
market category 9 (Large MPV)	-2.9783 (0.4176)	$[-3.8098 \; -2.1879]$
market category 10 (LUV)	-2.3235 (0.2707)	$[-2.8332 \; -1.7798]$
navigation	0.3209 (0.0612)	[0.2016 0.4399]
sun roof	0.1472 (0.0581)	[0.0362 0.2610]
year 2010	0.6117 (0.0791)	[0.4531 0.7636]
year 2011	0.7060 (0.0769)	[0.5512 0.8538]
year 2012	0.5201 (0.0862)	[0.3468 0.6827]
year 2013	1.0558 (0.0721)	[0.9193 1.1997]
ratio of driving in town/city	0.5811 (0.1056)	[0.3702 0.7826]
4 wheel drive	1.0011 (0.0936)	[0.8177 1.1828]
blind spot assistance	-0.6116 (0.1123)	$[-0.8321 \; -0.3961]$

Table 4.6: Estimates of the parameters β in Equation (4.1)

4.5 Conclusions

Both transportation policy advisors and transportation literature stress the importance of understanding what prevents individuals from making environmentally sounder transport related decisions. Focusing on the personal car market, sales of environmentally friendly (green) vehicles are limited despite their existence for many years. Many studies have attempted to explain the reasons for this by focusing on technological aspects of such vehicles. Only recently, authors started to look at attitudes and opinions that may also influence the choice. There exist studies that demonstrate that the choice of environmentally friendly vehicle depends on individual's environmental consciousness, with more environmentally conscious consumers opting for an environmentally friendly vehicle. However, there have been no studies that would investigate if and how this effect varies with individual characteristics, which is surprising given the consumer heterogeneity. In this chapter, we introduce a new way to model the effect of environmental consciousness on the choice of environmentally friendly vehicle as a function of a relevant individual characteristic. In our application, we choose this characteristic to be one's age. Because we do not know the corresponding function a-priori, instead of assuming some parametric shape and estimating the resulting parameters, we propose a model that infers this dependence from data. The only assumption that we make is that the environmental consciousness parameter varies smoothly with respect to individual's age. This means that the parameter measuring the impact of an individual's environmental consciousness on his or her vehicle choice is more similar for individuals that are of similar age. This smoothness is controlled by a single parameter which we also estimate.

We apply our approach to a revealed preference study of new vehicles' purchases in France in years $2010 - 2014$ and demonstrate the added value of it. Because of the revealed preference nature of our data set, we believe that the results presented are of importance in a real - environmentally friendly versus conventional - vehicle choice situation. In this application, we assume that the effect of environmental consciousness on the choice of an environmentally friendly vehicle depends on the individual's age. However, any other characteristic of interest could be chosen instead. Further, the model is not restricted to only latent environmental consciousness, meaning that other latent variables (such as latent prestige, for example) can be employed, too. Furthermore, the variables for which the parameters

are assumed to be functions of individual characteristics do not necessarily have to be latent either. Finally, the model is capable of incorporating more than one parameter which is assumed to be a function of some individual characteristic, where the covariance between these parameters could be then estimated as well.

We hope that this chapter serves as the first attempt in transportation literature to introduce flexible heterogeneous effects across the population in a choice model. Individuals are very different and therefore their sensitivities are expected to be different as well. Assuming them to be equal is too restrictive and does not allow to use the obtained results for personalizing incentives, which is very important if major goals, such as a cleaner environment, are to be achieved. In the future, it would be interesting to apply the same type of analysis to other countries and/or to include more latent attitude variables in the choice model.

4.A Sampling distributions

In this appendix, we derive the sampling distributions for all the model parameters. To deal with the binary probit model, we also sample the latent utilities in Equation (4.1). As we do not observe both the latent utilities and the latent environmental consciousness, we sample y_n^* and ENV_n alongside the model parameters. We therefore derive the sampling distributions for y_n^* and ENV_n as well.

We begin by rewriting Equation (4.3) as

$$I_n = \delta + \lambda ENV_n + \upsilon_n, \quad \upsilon_n \sim N(0_{K\times 1}, \Theta), \tag{4.6}$$

where $I_n = (I_{n1}, I_{n2}, \ldots, I_{nK})'$, $\delta = (\delta_{I_1}, \delta_{I_2}, \ldots, \delta_{I_K})'$, $\lambda = (\lambda_{I_1}, \lambda_{I_2}, \ldots, \lambda_{I_K})'$ and $\upsilon_n = (\upsilon_{n1}, \upsilon_{n2}, \ldots, \upsilon_{nK})'$. We denote $\Theta = \text{diag}(\sigma^2_{I_1}, \sigma^2_{I_2}, \ldots, \sigma^2_{I_K})$ the variance matrix for υ_n. In the first step, we sample the latent environmental consciousness, ENV_n, taking into account the information provided by the indicators. Note that we do not include the (indirect) information coming from y_n^*, see the discussion in Section 4.3.2. Considering Equations (4.4) and (4.6) jointly, we can write

$$\begin{pmatrix} ENV_n \\ I_n^* \end{pmatrix} \sim N\left(\begin{pmatrix} z_n'\alpha \\ \lambda^* z_n'\alpha \end{pmatrix}, \begin{pmatrix} 1 & \lambda^{*\prime} \\ \lambda^* & \lambda^*\lambda^{*\prime} + \mathbf{1}_{K\times K} \end{pmatrix} \right),$$

where $I_n^* = (\Theta^*)^{-1} I_n - (\Theta^*)^{-1}\delta$, $\lambda^* = (\Theta^*)^{-1}\lambda$ and Θ^* is the Cholesky factor of Θ. Here and further, $\mathbf{1}$ denotes an identity matrix with dimensions indicated in the subscript. Then, the conditional distribution of ENV_n, $\forall n$, is given by $ENV_n | I_n^*, \beta, \gamma(\overrightarrow{age}), g, \tau, \phi, \alpha, \delta, \lambda, \Theta \sim N(\mu_{ENV_n}, \sigma^2_{ENV})$ with

$$\mu_{ENV_n} = z_n'\alpha + \lambda^{*'}(\lambda^*\lambda^{*'} + \mathbf{1}_{K\times K})^{-1}(I_n^* - \lambda^* z_n'\alpha),$$
$$\sigma^2_{ENV} = 1 - \lambda^{*'}(\lambda^*\lambda^{*'} + \mathbf{1}_{K\times K})^{-1}\lambda^*.$$

Note that the expression for σ^2_{ENV} is independent of n.

Once we obtain a draw of ENV_n, $\forall n$, we can treat Equation (4.4) as a simple linear regression and sample α from the posterior $\alpha | y_n, I_n, ENV_n, y_n^*, \beta, \gamma(\overrightarrow{age}), g, \tau, \phi, \delta, \lambda, \Theta \sim N(\mu_\alpha, \sigma^2_\alpha)$ with $\mu_\alpha = (\sum_{n=1}^N z_n z_n' + \check{V}_\alpha^{-1})^{-1}(\sum_{n=1}^N z_n ENV_n + \check{V}_\alpha^{-1}\check{\alpha})$ and $\sigma^2_\alpha = (\sum_{n=1}^N z_n z_n' + \check{V}_\alpha^{-1})^{-1}$ where $p(\alpha) = N(\check{\alpha}, \check{V}_\alpha)$ is the prior distribution for α.

To sample δ and λ, denote $E_n = (1\ ENV_n)$ and rewrite Equation (4.6) as

$$I = (\mathbf{E} \otimes \mathbf{1}_{K\times K})\Lambda + \Upsilon, \quad \Upsilon \sim N(0_{NK\times 1}, \mathbf{1}_{N\times N} \otimes \Theta)$$

with $I = (I_1', I_2', \ldots I_N')'$, $\mathbf{E} = (E_1', E_2', \ldots, E_N')'$, $\Lambda = (\delta'\ \lambda')'$ and $\Upsilon = (\upsilon_1', \upsilon_2', \ldots, \upsilon_N')'$. Here and further, 0 denotes a zero vector with dimensions indicated in the subscript. We then assume the prior $\Lambda \sim N(\check{\Lambda}, \check{V}_\Lambda)$ and sample Λ from the multivariate normal distribution with mean $\left((\mathbf{E}'\mathbf{E} \otimes \Theta^{-1}) + \check{V}_\Lambda^{-1}\right)^{-1}\left(\mathrm{vec}(\Theta^{-1}(\mathbf{I}_{\mathrm{matrix}})'\mathbf{E}) + \check{V}_\Lambda^{-1}\check{\Lambda}\right)$ and variance $\left((\mathbf{E}'\mathbf{E} \otimes \Theta^{-1}) + \check{V}_\Lambda^{-1}\right)^{-1}$, where $\mathbf{I}_{\mathrm{matrix}}$ is a $N \times K$ matrix of the indicators such that $\mathrm{vec}\left((\mathbf{I}_{\mathrm{matrix}})'\right) = I$.

To sample the elements of the diagonal matrix Θ, we draw each of the $\sigma^2_{I_k}$ from an inverted Gamma-2 distribution with parameters $(\mathbf{I}^k_{\mathrm{matrix}} - \mathbf{E}\Lambda_k)'(\mathbf{I}^k_{\mathrm{matrix}} - \mathbf{E}\Lambda_k)$ and N degrees of freedom, where $\mathbf{I}^k_{\mathrm{matrix}}$ corresponds to the k^{th} column of $\mathbf{I}_{\mathrm{matrix}}$ and $\Lambda_k = (\delta_{I_k}\ \lambda_{I_k})'$.

In order to derive the sampling distributions for the remaining parameters, rewrite the utility in Equation (4.1) as

$$y_n^* = X_n\beta + (ENV_n i_n)\gamma(\overrightarrow{age}) + \varepsilon_n,$$

where i_n is a $1 \times r$ selection row vector indicating which of the observed age values corresponds to individual n. It selects the appropriate parameter from $\gamma(\overrightarrow{age})$. Stacking the

observations for all n gives

$$y^* = \mathbf{X}\beta + ENV^*\gamma(\overrightarrow{age}) + \varepsilon \tag{4.7}$$

with $y^* = (y_1^*, y_2^*, \ldots, y_N^*)'$, $\mathbf{X} = (X_1', X_2', \ldots, X_N')'$, $ENV^* = \big((ENV_1 i_1)', (ENV_2 i_2)', \ldots, (ENV_N i_N)'\big)'$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$.

To estimate the binary probit model, we sample the latent utilities y_n^*, $\forall n$, alongside the parameters in a Gibbs sampler as

$$y_n^* \big| y_n, I, ENV_n, \beta, \gamma(\overrightarrow{age}), g, \tau, \phi, \alpha, \delta, \lambda, \Theta \sim \begin{cases} N\big(X_n\beta + \gamma(age_n)ENV_n, 1\big)\mathcal{I}\big(y_n^* > 0\big) & \text{if } y_n = 1 \\ N\big(X_n\beta + \gamma(age_n)ENV_n, 1\big)\mathcal{I}\big(y_n^* \leq 0\big) & \text{if } y_n = 0 \end{cases}$$

with $\mathcal{I}(\cdot)$ denoting the indicator function. For details, see Albert and Chib (1993); Allenby and Rossi (1999); Train (2009); McCulloch and Rossi (1994, 2000).

More care is needed when sampling τ and $\gamma(\overrightarrow{age})$. The reason is that if $\tau \to \infty$ or if some of the observed age values are almost the same, the corresponding parameters in $\gamma(\overrightarrow{age})$ become extremely correlated, meaning that the correlation matrix $\mathbf{\Omega}$ becomes (nearly) singular and its traditional inverse does not exist. Therefore, to make the estimation numerically tractable when $\tau \to \infty$ or $age' \approx age''$, we use a singular value decomposition on $\mathbf{\Omega}$ and write

$$\gamma(\overrightarrow{age}) = g + fU\mathbf{\Omega}^{*\frac{1}{2}}\gamma^*(\overrightarrow{age}) \tag{4.8}$$

with $\gamma^*(\overrightarrow{age}) \sim N(0_{r^* \times 1}, \mathbf{1}_{r^* \times r^*})$, where $f = \sqrt{\phi}$ and U and $\mathbf{\Omega}^*$ ($\mathbf{\Omega}^* = \mathbf{\Omega}^{*\frac{1}{2}}\mathbf{\Omega}^{*\frac{1}{2}'}$) are obtained from the singular value decomposition of $\mathbf{\Omega}$, where we drop the singular values that are "almost zero" in $\mathbf{\Omega}^*$ and next drop the corresponding columns in U, leading to the (approximate) relationship $\mathbf{\Omega} \approx U_{r \times r^*}\mathbf{\Omega}^*_{r^* \times r^*}(U_{r \times r^*})'$, where r^* is the number of non-zero singular values. The threshold which we use to decide if a singular value is "almost zero" is 10^{-6}. This allows us to sample $\gamma^*(\overrightarrow{age})$ instead of $\gamma(\overrightarrow{age})$ directly, and next calculate $\gamma(\overrightarrow{age})$ using the expression in Equation (4.8). Note that the distribution of $\gamma^*(\overrightarrow{age})$ implies $\gamma(\overrightarrow{age}) \sim N(g, \phi\mathbf{\Omega})$, as we defined before. It is important to also note that $\gamma^*(\overrightarrow{age})$ is sampled in the necessarily lower dimensional space in case some of the correlations in $\mathbf{\Omega}$ become too large, that is, the dimension of $\gamma^*(\overrightarrow{age})$ is r^* with $r^* \leq r$. For details, see

Chapter 3. Following this discussion, rewrite the model in Equation (4.7) as

$$\begin{aligned} y^* &= \mathbf{X}\beta + ENV^*\big(g + FU\boldsymbol{\Omega}^{*\frac{1}{2}}\gamma^*(\overrightarrow{age})\big) + \varepsilon \\ &= \mathbf{X}\beta + \begin{pmatrix} ENV_1 \\ \vdots \\ ENV_N \end{pmatrix} g^* + ENV^*FU\boldsymbol{\Omega}^{*\frac{1}{2}}\gamma^*(\overrightarrow{age}) + \varepsilon \\ &= \mathbf{X}\beta + ENVg^* + ENV^*FU\boldsymbol{\Omega}^{*\frac{1}{2}}\gamma^*(\overrightarrow{age}) + \varepsilon \\ &= \mathbf{X}^*\beta^* + \varepsilon, \end{aligned} \tag{4.9}$$

where $ENV = (ENV_1, ENV_2, \ldots, ENV_N)'$, $\beta^* = \big(\beta'\ g^*\ \gamma^*(\overrightarrow{age})'\big)'$ and $\mathbf{X}^* = \big(\mathbf{X}'\ ENV'\ (ENV^*\ FU\boldsymbol{\Omega}^{*\frac{1}{2}})'\big)'$. The vector ENV contains the environmental consciousness value for each individual ignoring the fact that different individuals correspond to different age values. As priors, we assume $\beta \sim N(0, u\mathbf{1})$ and $g^* \sim N(0, u1)$, where we control for the prior uncertainty through the scalar u. Then, the prior for β^* becomes $\beta^* \sim N(0_{(d+1+r^*)\times 1}, \mathbf{B})$ with $\mathbf{B} = \operatorname{diag}(u\mathbf{1}_{(d+1)\times(d+1)}, \mathbf{1}_{r^*\times r^*})$, where we denote the dimension of β by d. Note that y^* becomes observable through the sampling step of y_n^*, $\forall n$, and, therefore, Equation (4.9) can be seen as a simple linear regression. We can now proceed to derive the posterior distributions of the remaining parameters.

First, note that we expect $\gamma(\overrightarrow{age})$ and τ to be (strongly) correlated, that is, if τ is large (small) we expect the parameters in $\gamma(\overrightarrow{age})$ to be rather similar (different) across the values of age and vice versa. Further, $\gamma(\overrightarrow{age})$ and g^* are also expected to be strongly correlated. Therefore, sampling all parameters from their full conditional posteriors is not efficient. To reduce the dependence between the draws, we sample β^* and τ jointly from $p\big(\beta^*\big|y_n, I, ENV_n, y^*, \tau, \phi, \alpha, \delta, \lambda, \Theta\big)$ and $p\big(\tau\big|y_n, I, ENV_n, y^*, \phi, \alpha, \delta, \lambda, \Theta\big)$, respectively, as

$p\big(\beta^*, \tau\big|y_n, I, ENV_n, y^*, \phi, \alpha, \delta, \lambda, \Theta\big) = p\big(\beta^*\big|y_n, I, ENV_n, y^*, \tau, \phi, \alpha, \delta, \lambda, \Theta\big) \times$
$p\big(\tau\big|y_n, I, ENV_n, y^*, \phi, \alpha, \delta, \lambda, \Theta\big)$. Note that the posterior distribution of τ is not conditional on $\gamma(\overrightarrow{age})$ and g^*. We then obtain the posterior distribution for τ as
$p\big(\tau\big|y_n, I, ENV_n, y^*, \phi, \alpha, \delta, \lambda, \Theta\big) \propto p\big(y^*\big|y_n, I, ENV_n, \tau, \phi, \alpha, \delta, \lambda, \Theta\big)p\big(\tau\big)$, where $p(\tau)$ is the prior distribution for τ, and where we integrate $p\big(y^*\big|y_n, I, ENV_n, \beta^*, \tau, \phi, \alpha, \delta, \lambda, \Theta\big)$

over β^* to obtain:

$$p\Big(y^*\Big|y_n, I, ENV_n, \tau, \phi, \alpha, \delta, \lambda, \Theta\Big) \propto \exp\Big(-\frac{1}{2}\Big(\mathbf{w}-\mathbf{V}\hat{\beta}\Big)'\Big(\mathbf{w}-\mathbf{V}\hat{\beta}\Big)\Big) \times \Big|\Big(\mathbf{V}'\mathbf{V}\Big)\Big|^{-\frac{1}{2}}$$

with $\mathbf{w} = \big(y^{*\prime} \quad (0_{(d+1+r^*)\times 1})'\big)'$, $\mathbf{V} = \big(\mathbf{X}^{*\prime} \quad (\mathbf{B}^{-\frac{1}{2}})'\big)'$ and $\hat{\beta} = (\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'\mathbf{w}$, where we write $\mathbf{B}^{-1} = \mathbf{B}^{-\frac{1}{2}'}\mathbf{B}^{-\frac{1}{2}}$. Since the obtained posterior distribution is not of a known type, a Metropolis-Hastings sampling step is needed for τ. We set the candidate generating function to be

$$\begin{aligned} \log(\tau^{\text{cand}}) &= \log(\tau^{\text{prev}}) + \eta, \\ \eta &\sim N(0, \kappa^2). \end{aligned} \tag{4.10}$$

The acceptance probability of τ^{cand} is then calculated as

$$\alpha = \min\left(\frac{p\big(\tau^{\text{cand}}\big|y_n, I, ENV_n, y^*, \phi, \alpha, \delta, \lambda, \Theta\big)g\big(\tau^{\text{prev}}\big|\tau^{\text{cand}}\big)}{p\big(\tau^{\text{prev}}\big|y_n, I, ENV_n, y^*, \phi, \alpha, \delta, \lambda, \Theta\big)g\big(\tau^{\text{cand}}\big|\tau^{\text{prev}}\big)}, 1\right), \tag{4.11}$$

where $g(\cdot)$ is the candidate generating function given in Equation (4.10) (Chib and Greenberg, 1995).

Next, we sample β^*. The kernel of the conditional posterior distribution for β^* is

$$\beta^*|y_n, I, ENV_n, y^*, \tau, \phi, \alpha, \delta, \lambda, \Theta \propto \exp\Big(-\frac{1}{2}\big(\beta^*-\bar{\beta}^*\big)'\big(\mathbf{X}^{*\prime}\mathbf{X}^* + (\mathbf{B})^{-1}\big)\big(\beta^*-\bar{\beta}^*\big)\Big)$$

with $\bar{\beta}^* = \big(\mathbf{X}^{*\prime}\mathbf{X}^* + (\mathbf{B})^{-1}\big)^{-1}\mathbf{X}^{*\prime}y^*$. This is a kernel of a multivariate normal distribution with mean $\bar{\beta}^*$ and covariance matrix $\big(\mathbf{X}^{*\prime}\mathbf{X}^* + (\mathbf{B})^{-1}\big)^{-1}$.

Finally, we sample ϕ from an inverted Gamma-2 ($IG2$) distribution with parameter $\big(\gamma(\overrightarrow{age}) - g\big)'\mathbf{\Omega}^{-1}\big(\gamma(\overrightarrow{age}) - g\big) + \mu$ and $r^* + \omega$ degrees of freedom, where we assume $p(\phi) \sim IG2(\mu, \omega)$ as a prior for ϕ.

We summarize our sampling procedure in Figure 4.7, where we graphically demonstrate Gibbs iterative sampling scheme from conditional posterior distributions.

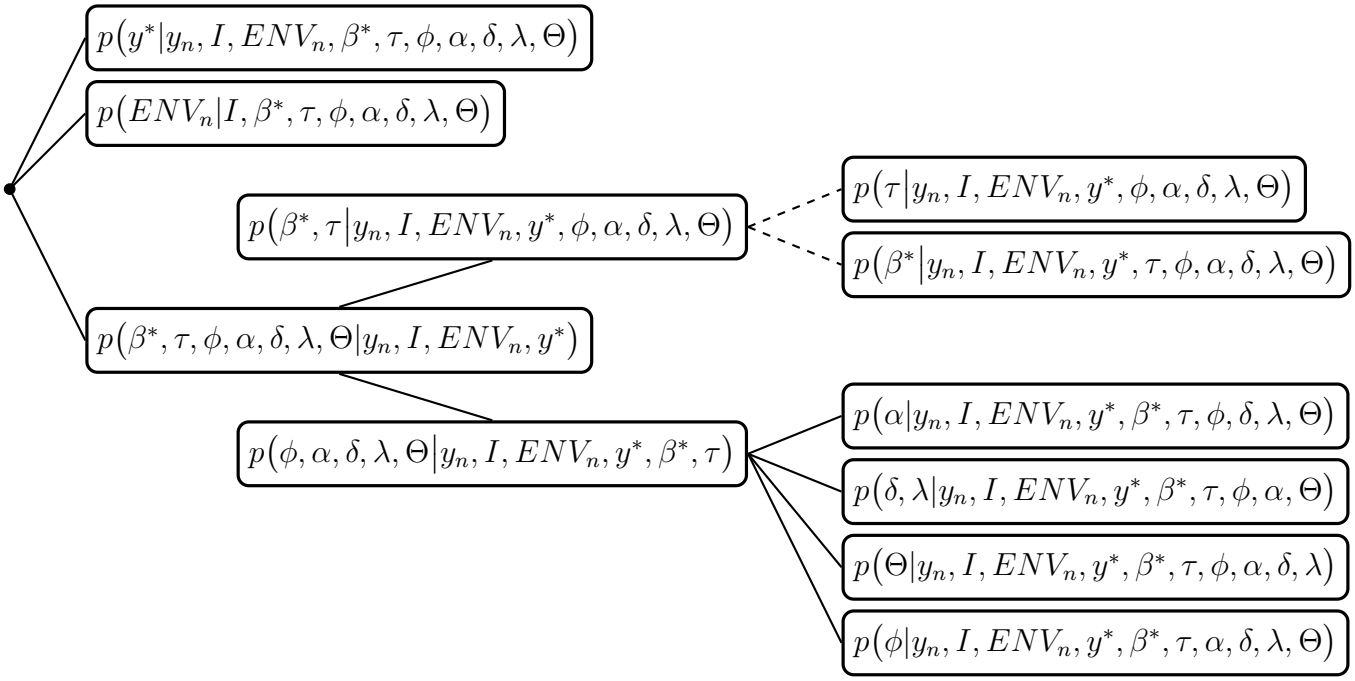


Figure 4.7: Scheme of the MCMC sampling steps. Dashed lines symbolize exact decompositions, solid lines symbolize decompositions based on Gibbs sampling

Chapter 5

Nederlandse Samenvatting (Summary in Dutch)

5.1 Inleiding

De verklarende variabelen bij sommige regressie-analyses zijn proporties van ingrediënten van een mengsel. Een kenmerk van deze proporties is dat zij sommeren tot 1. Bijvoorbeeld, yoghurt kan beschouwd worden als een mengsel van melk, fruit en stroop. In dit geval kan de term mengsel letterlijk genomen worden. In andere gevallen moet het mengsel figuurlijk geïnterpreteerd worden. Bijvoorbeeld, in een transportcontext kan de totale reistijd opgevat worden als een mengsel van wandeltijd, wachttijd en tijd in een trein of bus. In marketing kunnen advertentiecampagnes geïnterpreteerd worden als een mengsel van TV-spots, krantenadvertenties en advertenties in magazines. Ofschoon deze voorbeelden aantonen dat mengsels vrij frequent voorkomen, bestaan er diverse hiaten in de literatuur over mengseldata. Met deze dissertatie vul ik enkele van deze hiaten op.

Data voor regressie-analyses kunnen experimenteel of observationeel zijn. Omdat het verzamelen van experimentele data typisch omslachtig, tijdrovend of duur is, is het belangijk om de datacollectie zorgvuldig te plannen. In het jargon zeggen we dat een zorgvuldig gekozen proefopzet nodig is om efficiënt te experimenteren. Bij conjunctanalyses (*conjoint studies*) met mengsels is het gebruikelijk om respondenten te vragen om alternatieve mengsels te beoordelen door er een score (*rating*) aan toe te kennen, of door de mengsels te rangschikken (*ranking*) van minst tot meest geprefereerd. Deze aanpak is niet aan te be-

velen indien het aantal mengsels dat door één enkele respondent geëvalueerd moet worden groot is en de verschillen tussen de mengsel soms zeer subtiel zijn. In dergelijke scenarios worden keuze-experimenten (*choice experiments*) aanbevolen. In keuze-experimenten dienen respondenten herhaaldelijk een aantal alternatieve producten of diensten te beoordelen. Elke keer moeten zij hierbij het alternatief aanduiden dat ze prefereren. Tot aan de start van mijn doctoraatsonderzoek bestonden er geen technieken voor het optimaal plannen van keuze-experimenten met mengsels. Mijn dissertatie begint daarom met de ontwikkeling van methodologieën voor het construeren van optimale keuze-experimenten met mengsels. Hierbij veronderstel ik dat de voorkeur van de respondenten enkel afhangt van de proporties van de ingrediënten in het bestudeerde mengsel, maar niet van de totale hoeveelheid van het mengsel.

In sommige toepassingen zijn echter niet alleen de proporties van de ingrediënten bepalend voor de respons, maar ook de totale hoeveelheid van het mengsel. De respons van consumenten op advertentiecampagnes hangt bijvoorbeeld niet enkel af van de proportie advertenties op TV, de proportie advertenties in kranten en de proportie advertenties in magazines, maar natuurlijk ook van de totale hoeveelheid aan advertenties. Een regressiemodel voor het beschrijven van de afhankelijkheid tussen de uiteindelijke respons aan de ene kant en de proporties van de ingrediënten (advertenties in verschillende media) en de totale hoeveelheid aan de andere kant heet een *mixture-amount* model. De eerste mixture-amount modellen in de literatuur dateren van de jaren ’80 van de vorige eeuw. Deze modellen vertonen diverse tekortkomingen. Om deze reden ontwikkel ik in Hoofdstuk 3 van deze dissertatie een nieuw, flexibel model voor toepassingen waarin mengsels en totale hoeveelheden de verklarende variabelen zijn.

Hoofdstuk 4 is gebaseerd op een observationele studie over consumentenvoorkeuren voor elektrische en hybride voertuigen. Elektrische en hybride voertuigen zijn reeds geruime tijd beschikbaar op de markt. Echter, ondanks een aanzienlijke hoeveelheid promotie en stevige financiële stimuli is hun penetratie in de markt nog steeds zeer beperkt. Tal van onderzoekers in de transportsector zijn op zoek naar verklaringen voor deze beperkte marktpenetratie. Deze onderzoekers ontwikkelen geavanceerde wiskundige en statistische modellen en technieken voor een geschikte datacollectie teneinde de redenen voor de beperkte verspreiding van milieuvriendelijke voertuigen te identificeren. Er is bij al dit onderzoek een tendens om in toenemende mate rekening te houden met individuele kenmerken van consumenten, bij-

voorbeeld met de houding van de consumenten ten opzichte van het milieu. Het is immers een gegeven dat houdingen en opinies grondig verschillen van individu tot individu. Tot op heden werden evenwel geen modellen voorgesteld om bij het modelleren van voorkeuren voor elektrische en hybride wagens rekening te houden individuele houdingen en opinies en een heterogene impact van deze houdingen op de uiteindelijke voorkeuren. In Hoofdstuk 4 introduceer ik daarom een nieuw model dat de impact van heterogene houdingen ten opzichte van het milieu in rekening brengt bij het modelleren van voorkeuren voor elektrische of hybride voertuigen.

5.2 Overzicht

Deze dissertatie bevat drie hoofdstukken, die onafhankelijk van elkaar gelezen kunnen worden. In deze paragraaf beschrijf ik deze hoofdstukken in enig detail.

Hoofdstuk 2 is gebaseerd op Ruseckaite et al. (2016b). In dit hoofdstuk introduceer ik mengselmodellen in de literatuur over het modelleren van keuzes en ontwikkel ik een algoritme voor het construeren van ontwerpen voor keuze-experimenten met mengsels. Keuze-experimenten helpen om te bepalen hoe voorkeuren voor producten of diensten afhangen van combinaties van ingrediënten. Bij dergelijke experimenten worden de respondenten geconfronteerd met groepen van al dan niet hypothetische producten of diensten, en wordt aan hen gevraagd om het geprefereerde product of de geprefereerde dienst aan te duiden in elke groep. Omdat keuze-experimenten duur zijn om uit te voeren, is het belangrijk om ze zorgvuldig te plannen. Dit betekent dat we de experimenten zodanig moeten opzetten dat zij precieze schattingen opleveren van de voorkeuren. Tot bij de start van mijn doctoraatsonderzoek bestond er geen literatuur over het optimaal opzetten van keuze-experimenten met mengsels. Om deze reden stel ik in Hoofdstuk 2 een methodologie voor voor het optimaal ontwerp van keuze-experimenten met mengsels. Ik stel twee soorten algoritmes voor, een *coordinate-exchange* algoritme en een *particle swarm optimization* algoritme. Ik toon aan dat er een aanzienlijke toegevoegde waarde is in termen van statistische efficiëntie indien de resulterende experimentele ontwerpen gebruikt worden. Ter illustratie gebruik ik een mengselexperiment beschreven door Courcoux and Séménou (1997). Het doel van dit mengselexperiment was om de voorkeuren voor een specifieke cocktail te modelleren. Het ontwerp gehanteerd door Courcoux and Séménou (1997) is een ad-hoc ontwerp dat gebruik

maakt van zeven verschillende mengsels (= cocktails) van mangosap, cassis en limoensap. De mengsel verschillen in de proporties die gebruikt worden voor de drie ingrediënten. Aan het experiment namen 60 respondenten deel, die elk acht paren van cocktails dienden te proeven en hierbij telkens hun voorkeur kenbaar dienden te maken. Op het einde van Hoofdstuk 2 vergelijk ik het ad-hoc ontwerp van Courcoux and Séménou (1997) met een optimaal ontwerp dat ik zelf berekende met de eerder genoemde algoritmes.

In tal van scenario's hangt de bestudeerde responsvariabele enkel af van de proporties van de mengselingrediënten. In dergelijke scenario's volstaan mengselmodellen. In andere scenario's is de responsvariabele ook een functie van de totale hoeveelheid van het mengsel. In dergelijke gevallen dienen *mixture-amount* modellen aangewend te worden. Adverteerders moeten bijvoorbeeld beslissen wat het totale budget zal zijn voor een reclamecampagne en welke gedeeltes of proporties zij van dit budget zullen spenderen aan TV-advertenties, radioadvertenties en krantenadvertenties. Om de afhankelijkheid van een respons van proporties en van een totale hoeveelheid te modelleren worden momenteel de mengselparameters als een parametrische functie van de hoeveelheid gezien. De motivering voor deze aanpak is dat, indien de hoeveelheid de respons beïnvloedt, dit gebeurt door de mengselparameters te laten variëren. De zwakte van deze modellen is dat zij starten met een vooraf gespecificeerde functionele relatie tussen de mengselparameters en de hoeveelheid. Een correcte specificatie van deze functionele relatie is geen sinecure. Bovendien impliceert een flexibele specificatie een groot aantal modelparameters. In Hoofdstuk 3, dat gebaseerd is op Ruseckaite et al. (2016a), introduceer ik een nieuw type van dit model dat flexibel is en toch weinig parameters bevat. Het model maakt gebruik van zogenaamde Gaussiaanse processen en vereist geen a-priori specificatie van de functionele relatie tussen de mengselparameters en de hoeveelheid. Ik demonstreer tevens de toegevoegde waarde van het nieuwe model in vergelijking met bestaande modellen voor *mixture-amount* data. Ik gebruik hiervoor twee voorbeelden. Het eerste voorbeeld handelt over de reactie van muizen op een mengsel voor hormonen dat in verschillende dosissen wordt toegediend. Het tweede voorbeeld handelt over de herkenning van advertentiecampagnes. Het mengsel in dit tweede voorbeeld is een *media mix*, dit wil zeggen een combinaties van een proportie van TV-advertenties en een proportie aan advertenties in magazines. De hoeveelheid in dit voorbeeld is het totale beschikbare budget voor een advertentiecampagne. De responsvariabele is een binaire variabele die aangeeft of een bepaalde advertentiecampagne herkend wordt door de respondent of niet.

Hoofdstuk 4 is gebaseerd op een samenwerking met Michel Bierlaire gedurende een studieverblijf aan het TRANSP-OR laboratorium aan de École Polytechnique Fédérale de Lausanne (EPFL) in Zwitserland. Dit hoofdstuk maakt gebruik van een observationele dataset over aankopen van milieuvriendelijke wagens in Frankrijk in de periode 2010-2014. Voor deze data ontwikkel ik een nieuw keuzemodel dat latent milieubewustzijn in rekening brengt. Het model laat een heterogene impact van dit latente kenmerk toe op de consumentenvoorkeuren. In het model neem ik aan dat milieubewustzijn een functie is van leeftijd, maar ik had evengoed andere individuele consumentenkenmerken kunnen gebruiken. Omdat het a-priori onduidelijk is welk functionele relatie er is tussen de impact van milieubewustzijn en leeftijd maak ik opnieuw gebruik van Gaussiaanse processen. Uiteraard houdt het model ook rekening met technische kenmerken van voertuigen en met de prijs ervan. De klemtoon in Hoofdstuk 4 ligt echter op het begrijpen van de invloed van het milieubewustzijn op de voertuigkeuze. Ik illustreer de toegevoegde waarde van het nieuwe model in vergelijking met een model waarin de invloed van het milieubewustzijn voor alle individuele consumenten dezelfde is.

Bibliography

Agresti, A., 2002. Categorical Data Analysis. John Wiley & Sons, Inc.

Albert, J., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88, 669–679.

Aleksandrovs, L., Goos, P., Dens, N., De Pelsmacker, P., 2015. Mixed-media modeling may help optimize campaign recognition, brand interest: How to apply the "Mixture-amount modeling" method to cross-platform effectiveness measurement". Journal of Advertising Research 55, 443–457.

Allenby, G., Rossi, P., 1999. Marketing models of consumer heterogeneity. Journal of Econometrics 89, 57–78.

Archtnicht, M., Buhler, G., Hermeling, C., 2011. The impact of fuel availability on demand for alternative-fuel vehicles. Transportation Research Part D 17, 262–269.

Atkinson, A., Donev, A., Tobias, R., 2007. Optimum Experimental Designs, with SAS. Oxford Statistical Science Series.

Beck, M., Rose, J., Greaves, S., 2016. I can't believe your attitude: a joint estimation of best worst attitudes and electric vehicle choice. Transportation. doi: 10.1007/s11116-016-9675-9.

Beck, M., Rose, J., Hensher, D., 2013. Environmental attitudes and emissions charging: an example of policy implications for vehicle choice. Transportation Research Part A 50, 171–182.

Beggs, S., Cardell, S., 1980. Choice of smallest car by multi-vehicle households and the demand for electric vehicles. Transportation Research Part A 14, 389–404.

Beggs, S., Cardell, S., Hausman, J., 1981. Assessing the potential demand for electric cars. Journal of Econometrics 17, 1–19.

Ben-Akiva, M., Lerman, S., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D., Daly, A., de Palma, A., Gopinath, D., Karlstrom, A., Munizaga, M., 2002. Hybrid Choice Models: Progress and Challenges. Marketing Letters 13, 163–175.

Ben-Israel, A., Greville, T., 2003. Generalized Inverses: Theory and Applications. Springer.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer.

Bliemer, M., Rose, J., 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transportation Research Part B 44, 720–734.

Bliemer, M., Rose, J., Hensher, D., 2009. Efficient stated choice experiments for estimating nested logit models. Transportation Research Part B 43, 19–35.

Bliemer, M., Rose, J., Hess, S., 2008. Approximation of Bayesian efficiency in experimental choice designs. Journal of Choice Modeling 1, 98–127.

Bonilla, E., Chai, A., Williams, C., 2007. Multi-task Gaussian process prediction. In: Platt, J., Koller, D., Singer, Y., Roweis, S. (Eds.), Advances in Neural Information Processing Systems 20.

Boyle, P., Frean, M., 2005. Dependent Gaussian processes. In: Saul, L., Weiss, Y., Bottou, L. (Eds.), Advances in Neural Information Processing Systems 17. The MIT Press, pp. 217–224.

Brent, R., 1973. Algorithms for Minimization without Derivatives. Prentice-Hall.

Brownstone, D., Bunch, D., Golob, T., Ren, W., 1996. A transactions choice model for forecasting demand for alternative-fuel vehicles. Research in Transportation Economics 4, 87–129.

Brownstone, D., Bunch, D., Train, K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. Transportation Research Part B 34, 315–338.

Brownstone, D., Train, K., 1999. Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics 89, 109–129.

Bunch, D., Bradley, M., Golob, T., Kitamura, R., Occhiuzzo, G., 1993. Demand for clean-fuel vehicles in California: A discrete choice stated preference pilot project. Transportation Research Part A 27, 237–253.

Burgess, L., Street, D., 2005. Optimal designs for choice experiments with asymmetric attributes. Journal of Statistical Planning and Inference 134, 288–301.

Casella, G., George, E., 1992. Explaining the Gibbs sampler. The American Statistician 46, 167–174.

Chen, R., Chang, S., Wang, W., Wong, W., 2011. Optimal experimental designs via particle swarm optimization methods, preprint.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. The American Statistician 49, 327–335.

Claringbold, P., 1955. Use of the simplex design in the study of the joint action of related hormones. Biometrics 11, 174–185.

Clerc, M., Kennedy, J., 2002. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. IEEE Transactions on Evolutionary Computation 6, 58–73.

Cornell, J., 2002. Experiments with Mixtures. John Wiley & Sons, Inc.

Courcoux, P., Séménou, M., 1997. Une méthode de segmentation pour l'analyse de données issues de comparaisons par paires. Revue de Statistique Appliquée 45, 59–69.

Danthurebandara, V., Yu, J., Vandebroek, M., 2011. Effect of choice complexity on design efficiency in conjoint choice experiments. Journal of Statistical Planning and Inference 141, 2276–2286.

David, H., 1963. The Method of Paired Comparisons. Charles Griffin & Co.

Daziano, R., 2010. A Bayesian approach to hybrid choice models. Ph.D. thesis, Université Laval Quebec.

Daziano, R., Bolduc, D., 2013. Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. Transportmetrica A 9, 74–106.

De Pelsmacker, P., Geuens, M., Van Den Bergh, J., 2010. Marketing Communications: A European Perspective. Financial Times Management.

Duvenaud, D., Lloyd, J., Grosse, R., Tenenbaum, J., Ghahramani, Z., 2013. Structure discovery in nonparametric regression through compositional kernel search. Proceedings of the 30th International Conference on Machine Learning, 1166–1174.

European commission, 2016. Transport: Electric vehicles. `http://ec.europa.eu/transport/themes/urban/vehicles/road/electric_en.htm`, [Online; accessed 10 March, 2016].

Franses, P., Paap, R., 2001. Quantitative Models in Marketing Research. Cambridge University Press, Cambridge.

Gattiker, J., Hamada, M., Higdon, D., Schonlau, M., Welch, W., 2015. Using a Gaussian process as a nonparametric regression model. Quality and Reliability Engineering International 32, 673–680.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 515–533.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D., 2013. Bayesian Data Analysis. Chapman and Hall/CRC.

Gilmour, S., Goos, P., 2009. Analysis of data from non-orthogonal multistratum designs in industrial experiments. Journal of the Royal Statistical Society, Ser. C 58, 467–484.

Glerum, A., Bierlaire, M., 2012. Accounting for response behavior heterogeneity in the measurement of attitudes: an application to demand for electric vehicles. Proceedings of the Swiss Transport Research Conference (STRC).

Glerum, A., Stakovikj, L., Thémans, M., Bierlaire, M., 2013. Forecasting the demand for electric vehicles: accounting for attitudes and perceptions. Transportation Science 48, 483–499.

Goos, P., 2002. The Optimal Design of Blocked and Split-Plot Experiments. Springer.

Goos, P., Donev, A., 2007. D-optimal minimum support mixture designs in blocks. Metrika 65, 53–68.

Goos, P., Vermeulen, B., Vandebroek, M., 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Journal of Statistical Planning and Inference 140, 851–861.

Grasshoff, U., Grossmann, H., Holling, H., Schwabe, R., 2003. Optimal paired comparison designs for first-order interactions. Statistics 37, 373–386.

Grasshoff, U., Grossmann, H., Holling, H., Schwabe, R., 2004. Optimal designs for main effects in linear paired comparison models. Journal of Statistical Planning and Inference 126, 361–376.

Greenberg, E., 2014. Introduction to Bayesian Econometrics. Cambridge University Press.

Grimm, P., 2010. Social Desirability Bias. Wiley International Encyclopedia of Marketing. 2.

Grossmann, H., Grasshoff, U., Schwabe, R., 2009. Approximate and exact optimal designs for paired comparisons of partial profiles when there are two groups of factors. Journal of Statistical Planning and Inference 139, 1171–1179.

Grossmann, H., Holling, H., Grasshoff, U., Schwabe, R., 2006. Optimal designs for asymmetric linear paired comparisons with a profile strength constraint. Metrika 64, 109–119.

Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Matematik 2, 84–90.

Hensher, D., Rose, J., 2009. Simplifying choice through attribute preservation or non-attendance: Implications for willingness to pay. Transportation Research Part E 45, 583–590.

Hensher, D., Rose, J., Greene, W., 2005. Applied Choice Analysis: A Primer. Cambridge University Press.

Hess, S., Beharry-Borg, N., 2012. Accounting for latent attitudes in willingness-to-pay studies: The case of coastal water quality improvements in Tobago. Environmental and Resource Economics 52, 109–131.

Hess, S., Shires, J., Jopson, A., 2013. Accommodating underlying pro-environmental attitudes in a rail travel context: application of a latent variable latent class specification. Transportation Research Part D 25, 42–48.

Hjorthol, R., 2013. Attitudes, ownership and use of electric vehicles – a review of literature. TOI report 1261/2013, Institute of Transport Economics. Norvegian Centre for Transport Research.

Huber, J., Zwerina, K., 1996. The importance of utility balance in efficient choice designs. Journal of Marketing Research 33, 307–317.

Hurtubia, R., Nguyen, M., Glerum, A., Bierlaire, M., 2014. Integrating psychometric indicators in latent class choice models. Transportation Research Part A 64, 135–146.

Johansson, M., Heldt, T., Johansson, P., 2005. Latent variables in a travel mode choice model: Attitudinal and behavioural indicator variables. Working paper, Uppsala University, Department of Economics.

Kennedy, J., 1997. The particle swarm: Social adaptation of knowledge. IEEE International Conference on Evolutionary Computation, 303–308.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. IEEE International Conference on Neural Networks 4, 1942–1948.

Kessels, R., Jones, B., Goos, P., 2011a. Bayesian optimal designs for discrete choice experiments with partial profiles. Journal of Choice Modelling 4, 52–74.

Kessels, R., Jones, B., Goos, P., Vandebroek, M., 2009. An efficient algorithm for constructing Bayesian optimal choice designs. Journal of Business & Economic Statistics 27, 279–291.

Kessels, R., Jones, B., Goos, P., Vandebroek, M., 2011b. The usefulness of Bayesian optimal designs for discrete choice experiments. Applied Stochastic Models in Business and Industry 27, 173–188.

Kiefer, J., 1961. Optimum design in regression problems II. Annals of Mathematical Statistics 32, 298–325.

Lawrence, N., 2004. Gaussian process latent variable models for visualisation of high dimensional data. Advances in Neural Information Processing Systems 16, 329–336.

Leonard, T., 1978. Density estimation, stochastic processes and prior information. Journal of the Royal Statistical Society, Ser. B 40, 113–146.

McCulloch, R., Rossi, P., 1994. An exact likelihood analysis of the multinomial probit model. Journal of Econometrics 64, 207–240.

McCulloch, R., Rossi, P., 2000. Bayesian analysis of the multinomial probit model. In: Mariano, R., Schuermann, T., Weeks, M. (Eds.), Simulation-Based Inference in Econometrics. Cambridge University Press, New York.

Melkumyan, A., Ramos, F., 2011. Multi-kernel Gaussian processes. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence 2, 1408–1413.

Meyer, R., Nachtsheim, C., 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. Technometrics 37, 60–69.

Neal, R., 1997. Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification. Technical Report No. 9702, Dept. of Statistics, University of Toronto.

Neal, R., 1999. Regression and classification using Gaussian process priors. In: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), Bayesian Statistics 6. Oxford University Press, pp. 475–501.

Piepel, G., 1982. Measuring component effects in constrained mixture experiments. Technometrics 24, 29–39.

Piepel, G., Cooley, S., Jones, B., 2005. Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm. Quality Engineering 17, 579–594.

Piepel, G., Cornell, J., 1985. Models for mixture experiments when the response depends on the total amount. Technometrics 27, 219–227.

Potoglou, D., Kanaroglou, P., 2007. Household demand and willingness to pay for clean vehicles. Transportation Research Part D 12, 264–274.

Rasmussen, C., Williams, C., 2005. Gaussian Processes for Machine Learning. The MIT Press.

Rehman, S., Paterson, A., Piggott, J., 2007. Optimisation of flours for chapatti preparation using a mixture design. Journal of the Science of Food and Agriculture 87, 425–430.

Riihimäki, J., Vehtari, A., 2014. Laplace approximation for logistic Gaussian process density estimation and regression. Bayesian Analysis 9, 425–448.

Robert, C., Casella, G., 2010. Monte Carlo Statistical Methods. Springer.

Rose, J., Bliemer, M., 2009. Constructing efficient stated choice experimental designs. Transport Reviews: A Transnational Transdisciplinary Journal 29, 587–617.

Ruseckaite, A., Fok, D., Goos, P., 2016a. Flexible mixture-amount models for business and industry using Gaussian processes, Tinbergen Institute Discussion Paper 16-075/III.

Ruseckaite, A., Goos, P., Fok, D., 2016b. Bayesian D-optimal choice designs for mixtures. Journal of the Royal Statistical Society, Ser. C. doi: 10.1111/rssc.12174.

Sahrmann, H., Piepel, G., Cornell, J., 1987. In search of the optimum Harvey Wallbanger recipe via mixture experiment techniques. The American Statistician 41, 190–194.

Salimans, T., 2012. Variable selection and functional form uncertainty in cross-country growth regressions. Journal of Econometrics 171, 267–280.

Sánchez, M., Lafuente, R., 2010. Defining and measuring environmental consciousness. Revista Internacional de Sociología 68, 731–755.

Sándor, Z., Wedel, M., 2001. Designing conjoint choice experiments using managers' prior beliefs. Journal of Marketing Research 38, 430–444.

Sándor, Z., Wedel, M., 2005. Heterogeneous conjoint choice designs. Journal of Marketing Research 42, 210–218.

Scheffé, H., 1958. Experiments with mixtures. Journal of the Royal Statistical Society, Ser. B 20, 344–360.

Scheffé, H., 1963. The simplex-centroid design for experiments with mixtures. Journal of the Royal Statistical Society, Ser. B 25, 235–263.

Shi, Y., Eberhart, R., 1998. A modified particle swarm optimizer. Proceedings of Congress on Evolutionary Computation, 69–73.

Smith, W., 2005. Experimental Design for Formulation. Siam.

Spanier, J., Maize, E., 1994. Quasi-random methods for estimating integrals using relatively small samples. SIAM Review 36, 18–44.

Street, D., Burgess, L., 2007. The Construction of Optimal Stated Choice Experiments: Theory and Methods. Wiley.

Tanaka, M., Takanori, I., Murakami, K., Friedman, L., 2014. Consumers' willingness to pay for alternative fuel vehicles: A comparative discrete choice analysis between the U.S. and Japan. Transportation Research Part A 70, 194–209.

Tanner, M., Wong, W., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–540.

Temme, D., Paulssen, M., Dannewald, T., 2008. Incorporating latent variables into discrete choice models - a simultaneous estimation approach using SEM software. Business Research 1, 220–237.

Train, K., 1980. The potential market for non-gasoline-powered automobiles. Transportation Research Part A 14, 405–414.

Train, K., 2009. Discrete choice methods with simulation. Cambridge University Press.

Uranisi, H., 1964. Optimal design for the special cubic regression model on the q-simplex. Mathematical Report 1, Kyushu University, General Education Department.

Vermeulen, B., Goos, P., Scarpa, R., Vandebroek, M., 2011. Bayesian conjoint choice designs for measuring willingness to pay. Environmental and Resource Economics 48, 129–149.

Walker, J., 2001. Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge.

Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. Mathematical Social Sciences 43, 303–343.

Wang, J., Fleet, D., Hertzmann, A., 2008. Gaussian process dynamical models for human motion. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 283–298.

Williams, C., 1998. Prediction with Gaussian processes: From linear regression to linear prediction and beyond. In: Jordan, M. (Ed.), Learning in Graphical Models. Vol. 89. Springer Netherlands, pp. 599–621.

Wilson, A., Adams, R., 2013. Gaussian process kernels for pattern discovery and extrapolation. In: Dasgupta, S., McAllester, D. (Eds.), Proceedings of The 30th International Conference on Machine Learning. Vol. 28. pp. 1067–1075.

Wong, W., Chen, R., Huang, C., Wang, W., 2015. A modified particle swarm optimization technique for finding optimal designs for mixture models. PLoS ONE 10. doi: 10.1371/journal.pone.0124720.

Yu, J., Goos, P., Vandebroek, M., 2009. Efficient conjoint choice designs in the presence of respondent heterogeneity. Marketing Science 28, 122–135.

Yu, J., Goos, P., Vandebroek, M., 2010. Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments. Transportation Research Part B 44, 1268–1289.

Yu, J., Goos, P., Vandebroek, M., 2011. Individually adapted sequential Bayesian designs for conjoint choice experiments. International Journal of Research in Marketing 28, 378–388.

Zellner, A., 1996. An Introduction to Bayesian Inference in Econometrics. Wiley-Interscience.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

619. A.G. KOPÁNYI-PEUKER, *Endogeneity Matters: Essays on Cooperation and Coordination*

620. X. WANG, *Time Varying Risk Premium and Limited Participation in Financial Markets*

621. L.A. GORNICKA, *Regulating Financial Markets: Costs and Trade-offs*

622. A. KAMM, *Political Actors playing games: Theory and Experiments*

623. S. VAN DEN HAUWE, *Topics in Applied Macroeconometrics*

624. F.U. BRÄUNING, *Interbank Lending Relationships, Financial Crises and Monetary Policy*

625. J.J. DE VRIES, *Estimation of Alonso's Theory of Movements for Commuting*

626. M. POPŁAWSKA, *Essays on Insurance and Health Economics*

627. X. CAI, *Essays in Labor and Product Market Search*

628. L. ZHAO, *Making Real Options Credible: Incomplete Markets, Dynamics, and Model Ambiguity*

629. K. BEL, *Multivariate Extensions to Discrete Choice Modeling*

630. Y. ZENG, *Topics in Trans-boundary River sharing Problems and Economic Theory*

631. M.G. WEBER, *Behavioral Economics and the Public Sector*

632. E. CZIBOR, *Heterogeneity in Response to Incentives: Evidence from Field Data*

633. A. JUODIS, *Essays in Panel Data Modelling*

634. F. ZHOU, *Essays on Mismeasurement and Misallocation on Transition Economies*

635. P. MULLER, *Labor Market Policies and Job Search*

636. N. KETEL, *Empirical Studies in Labor and Education Economics*

637. T.E. YENILMEZ, *Three Essays in International Trade and Development*

638. L.P. DE BRUIJN, *Essays on Forecasting and Latent Values*

639. S. VRIEND, *Profiling, Auditing and Public Policy: Applications in Labor and Health Economics*

640. M.L. ERGUN, *Fat Tails in Financial Markets*

641. T. HOMAR, *Intervention in Systemic Banking Crises*

642. R. LIT, *Time Varying Parameter Models for Discrete Valued Time Series*

643. R.H. KLEIJN, *Essays on Bayesian Model Averaging using Economic Time Series*

644. S. MUNS, *Essays on Systemic Risk*

645. B.M. SADABA, *Essays on the Empirics of International Financial Markets*

646. H. KOC, *Essays on Preventive Care and Health Behaviors*

647. V.V.M. MISHEVA, *The Long Run Effects of a Bad Start*

648. W. LI, *Essays on Empirical Monetary Policy*

649. J.P. HUANG, *Topics on Social and Economic Networks*

650. K.A. RYSZKA, *Resource Extraction and the Green Paradox: Accounting for Political Economy Issues and Climate Policies in a Heterogeneous World*

651. J.R. ZWEERINK, *Retirement Decisions, Job Loss and Mortality*

652. M. K. KAGAN, *Issues in Climate Change Economics: Uncertainty, Renewable Energy Innovation and Fossil Fuel Scarcity*

653. T.V. WANG, *The Rich Domain of Decision Making Explored: The Non-Triviality of the Choosing Process*

654. D.A.R. BONAM, *The Curse of Sovereign Debt and Implications for Fiscal Policy*

655. Z. SHARIF, *Essays on Strategic Communication*

656. B. RAVESTEIJN, *Measuring the Impact of Public Policies on Socioeconomic Disparities in Health*

657. M. KOUDSTAAL, *Common Wisdom versus Facts; How Entrepreneurs Differ in Their Behavioral Traits From Others*

658. N. PETER, *Essays in Empirical Microeconomics*

659. Z. WANG, *People on the Move: Barriers of Culture, Networks, and Language*

660. Z. HUANG, *Decision Making under Uncertainty-An Investigation from Economic and Psychological Perspective*

661. J. CIZEL, *Essays in Credit Risk, Banking, and Financial Regulation*

662. I. MIKOLAJUN, *Empirical Essays in International Economics*

663. J. BAKENS, *Economic Impacts of Immigrants and Ethnic Diversity on Cities*

664. I. BARRA, *Bayesian Analysis of Latent Variable Models in Finance*

665. S. OZTURK, *Price Discovery and Liquidity in the High Frequency World*

666. J. JI, *Three Essays in Empirical Finance*

667. H. SCHMITTDIEL, *Paid to Quit, Cheat, and Confess*

668. A. DIMITROPOULOS, *Low Emission Vehicles: Consumer Demand and Fiscal Policy*

669. G.H. VAN HEUVELEN, *Export Prices, Trade Dynamics and Economic Development*