THE EFFECT OF VISUALIZATION OF CHOICE OPTIONS ON ACTUAL CHOICE

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science Econometrics and Management Science
Specialization: Quantitative Marketing

to the

Erasmus University Rotterdam

By

Name:	C.N. Edelenbosch
Enrolment No:	296450
Supervisor:	Prof. Dr. P. Goos
Co-reader:	A. Ruseckaite

ERASMUS SCHOOL OF ECONOMICS
Department of Econometrics
Erasmus University Rotterdam

January 2014

Abstract

Consumers have to choose among a great variety of products with different attributes such as brand, design and price. Companies are eager to find out why consumers choose their product or why they do not. The product choice can depend on the product's characteristics, but it can also depend on the way the attribute levels are visualized. A clear understanding of what drives the consumer's product choices can lead to several improvements, resulting in higher sales and market shares. The goal of this thesis is to investigate how different visualization approaches affect consumer choice behavior.

The choice behavior is investigated using a choice-based conjoint (CBC) study for the choice of a credit card. Each respondent has to decide which of the product concepts shown in several successive choice tasks he prefers. Because people can react differently to a company's specific credit card offering, individual-level "part-worth" utilities are allowed for each respondent. A utility is a value representing the attractiveness of each feature in a conjoint study. Individual-level part-worth utilities allow for easy segmentation as they provide a way to detect different groups of respondents. This study uses the hierarchical Bayes (HB) estimation because of its ability to provide reasonable estimates for the utilities, based on only a few choices by each respondent. The software that is used for the estimation procedure is Sawtooth's implementation of CBC HB. To verify the results produced by Sawtooth, a HB estimation procedure has been implemented in the open source language R as well.

In this thesis, it is proven that visualization techniques have an effect on consumer choice behavior. Therefore, they are included in the design of market research exercises in order to increase realism, external validity, and to provide recommendations to optimally communicate a product line-up. The results were presented at the 23rd AMAs Annual Advanced Research Techniques (ART) Forum on June 25, 2012 in Seattle (WA, USA).

The content of this thesis is based on an internship at the SKIM, an international market research company specialized in conjoint analysis. The internship was jointly supervised by the Chief Methodology Officer G. Loosschilder, Research Director K. van der Wagt, Project Manager C. Borghi and Prof. Dr. P. Goos.

Keywords: choice-based conjoint, Hierarchical Bayes, Mixed logit model, Monte Carlo Markov Chain algorithm, Multinomial logit model, Respondent heterogeneity, Visualization effects

Table of Contents

List of Figures

List of Tables

1. **Introduction**

1.1 Conjoint methodologies

Market research is the discipline of identifying and analyzing the market need, market size, and competition. The goal is to acquire valuable information that can be used in taking various marketing decisions, such as operational or strategic decisions. The scope of market research is very wide and there are many of different analytical techniques that can be used, depending on the type of decision to be taken. For example, Apple may be interested in forecasting what the market for the iPhone looks like in the coming years. What type of customers will buy the next version of the iPhone? How would people respond to a similar product being launched on the market by the competitors? Why do people prefer Apple? These are all examples of questions that fall within the scope of market research. Different types of questions require different techniques to formulate answers.

In general, market research can be divided into two categories: qualitative and quantitative market research. Qualitative market research is the kind of market research that is focused on getting inside the consumers' minds. It is about finding out what people think and why they think that, in order to understand their feelings and motivations. The results are given in words or pictures, and are regularly presented by graphs. Quantitative market research is based on numerical analyses and statistical models, which are often presented by tables. A quantitative market research study often has over a thousand respondents, as this leads to more statistical power. A specific sample has to be chosen that resembles the population that is studied. Researchers use questionnaires and surveys to collect the required data. In contrary, qualitative market research has fewer respondents, but researchers often know more details about them. Examples of data gathering strategies are individual in-depth interviews and focus groups. Qualitative market research is ideal for the first phases of a research project, because its purpose is usually more explorative. Quantitative market research is used to test specific hypotheses. It results in significant proof, which leads to the acceptation or rejection of the hypotheses. Yet, despite the many differences between both types of market research, the ultimate goal is the same: studying human behavior to get a clear understanding of the market and to make good decisions based on the research questions (Guba and Lincoln, 1994).

Conjoint methodologies belong to the category of quantitative market research. They are based on direct data only. Historical data in theory could be applied as well, but that kind of data usually does not give enough relevant information. Therefore, data is collected for a particular study. This can be based on experiments, in which respondents have to complete several choice tasks that will reveal as much as possible concerning their buying behavior. Conjoint methodologies originated in the field of psychometrics. This is the branch of psychology that deals with the design and interpretation of quantitative tests for the measurement of certain psychological variables, such as knowledge, personality traits and intelligence (Luce and Tukey, 1964). The techniques from psychometrics were first applied to marketing by Green and Rao in 1971. Since then, conjoint methodologies have become a popular marketing research tool and excellent updates have been published that describe the more recent developments, for example, see Green and Srinivasan (1978, 1990).

Conjoint analysis *considers* the *joint* influence of a product's characteristics on the purchase decision. The products shown in a questionnaire are usually described by a combination of different attributes. Traditionally, respondents were asked to consider attributes one by one and evaluate them (Thurstone, 1927). For example, one was asked to indicate how important the screen width of the iPhone is, or the number of gigabytes. This is not the case for conjoint analysis, because with this type of analysis, respondents only state preferences for or give ratings about full products. A regression model is used, which consists of independent variables corresponding to the possible levels of all the attributes. The utilities that correspond to these variables are called "part-worths". By estimating these "part-worths", one can still retrieve the preferences for single attributes. One can also quantify the kinds of trade-offs respondents make.

Market researchers often want to predict what happens when a new product enters the market. Predictions about the future sales of a new product cannot be made by studying historical data and past trends in combination with extrapolation, because no historical data is available on new products. In case of a product upgrade, the historical data of the "older" product can be used, but it is not likely that this results in accurate predictions. For example, Apple is interested in the future sales of the iPhone 5. Compared to the latest version, the White iPhone 4, the iPhone 5 has a 4G network connection and a wider screen and so forth. The past trends can be used to predict the future sales (see figure 1.1), but it is rather difficult to get accurate estimates.

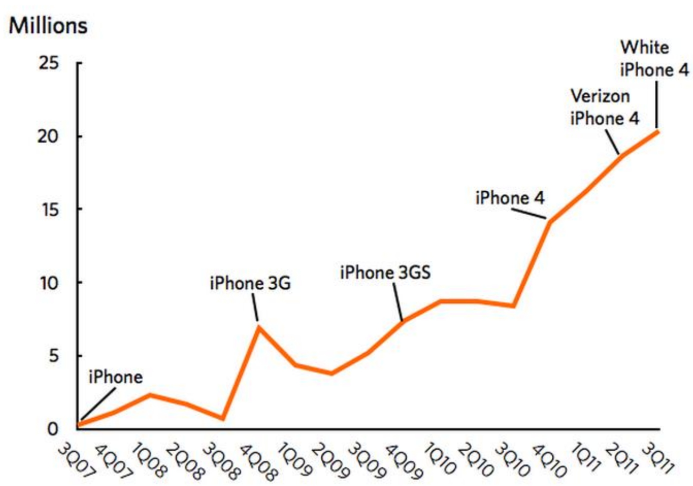


Figure 1.1: iPhone unit sales trend

Conjoint analysis makes it possible to study scenarios in which completely new or upgraded products are introduced. This technique is based on the fact that data is gathered for the specific need of the study. The full products consist of attributes that are already present on the market and attributes which might be introduced in the near future. This way, market researchers can estimate future market shares by using conjoint methodologies. This is a selling point which explains why conjoint analysis can be very useful for companies. Furthermore, conjoint techniques have the advantage of being realistic. People only consider a few products and compare them. The trade-off between these products is not based on single attributes only, it is rather based on a combination of attributes.

Conjoint analysis has many substantive applications concerning choice behavior and it is applied in various areas like consumer packaged goods (Desarbo *et al.*, 1995), durable goods (Wittink *et al.*, 1994), and healthcare (Ryan *et al.*, 1998). In practice, market research questions require measuring price elasticity, finding out which product characteristics individuals consider as important, as well as how much they want to pay for those characteristics and products. In addition to answering this type of questions, a powerful market simulator can be constructed to test different scenarios for a company's own and a competitor's actions and make market predictions to optimize profits and/or market shares.

Choice-based conjoint (CBC) is the most popular conjoint-related technique in use today in market research. In CBC experiments, respondents are shown multiple product concepts and asked to choose the one they would buy. The task that the respondents are asked to perform is similar to the actual decision they make when making a purchase. Thus, CBC experiments mimic real choice behavior, which is the main advantage of this technique. The goal of CBC studies is to estimate preferences respondents have for various levels of the different attributes. These preferences are described as numerical values, forming a set of utilities. These utilities represent the level of satisfaction a person receives from a certain level of the product attribute. The higher the utility, the higher the satisfaction. When making real-life choices, people tend to be rational and therefore maximize their total utility when making a choice.

The utilities present the importance of the levels of the attributes. The utilities are used to obtain an accurate estimate for the probability that a decision maker chooses a certain product concept. A multinomial logit model is used when the utilities for a level are assumed to be the same among the respondents. The multinomial logit model is described in detail in chapter 3. One can also allow for heterogeneity, which leads to different utilities across the respondents for the levels of the attributes. The hierarchical Bayes estimation procedure is then used, which is based on a multi-layered model. In this thesis, the focus is on the CBC HB estimation procedure for the mixed logit model, which is described in detail in chapter 4.

Standard CBC estimation procedures describe consumer preferences as a function of the products' attributes and attribute levels, but not of the way the attributes or their levels are visualized. In reality, however, product choices are impacted by the way attribute levels are presented. Visualization can make certain attributes or their levels more attractive. This can lead to a significant improvement in preference for the product. For example, the advertisement of a credit card from ING in figure 1.2 shows annual percentage yields at various account balance amounts. It seems like every dollar earns interest at 5.30%, which is very attractive. However, that is only true if your balance is $100,000 or more. If your balance is between $50,000 and $100,000 you earn only 5.05%; you earn 3.00% otherwise.

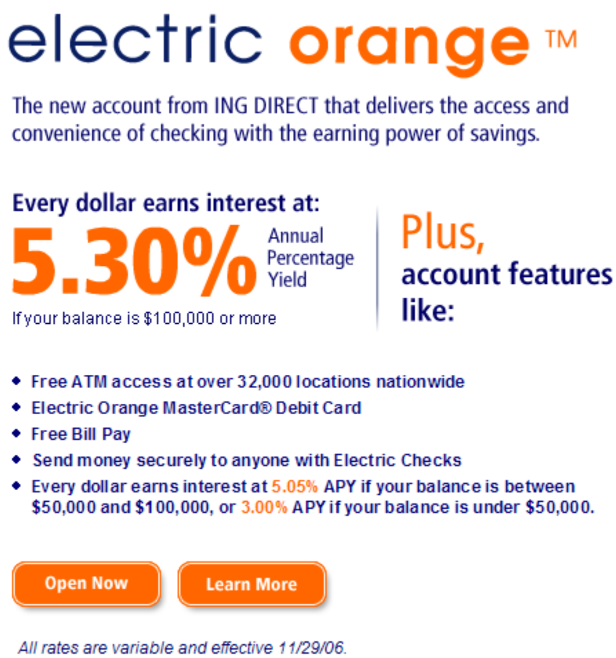


Figure 1.2: Advertisement for a credit card offer from ING

In the advertisement, the higher annual percentage yield is displayed more prominently, which makes the account more attractive. However, most of us keep a lower account balance, rendering the advertisement "misleading". The font size may influence utilities, and the yield at a high balance could have had a different value if the font size had been smaller. The same applies to other aspects of the offer. Product managers will fine-tune the way they present their products by giving the most favorable features the largest visual impact. Conversely, playing with visual impact is a way to communicate relevant information while avoiding clutter and overwhelming consumers. This can be translated into a need for a way to simultaneously estimate attribute importance and their visual impact, to optimize what to say and how to say it (Mueller *et al.*, 2010).

It is uncommon to study variations on the layout of textual stimuli. Myung (2003) and Feldman (2003) use conjoint analysis to optimally combine layout alternatives (font type, font size, line spacing, menu orientation etc.) in web environments. Unfortunately, both consider layout effects in “isolation”, only to find the optimal layout regardless of product characteristics. The advertisement in figure 1.2 implies that they can uncover the ideal font size for every level of the attribute “interest rate”; not which font size is more effective for different products at different interest rates.

Another interesting study to test the impact of visualization is done by Meiβner *et al.* (2010). They used eye tracking to examine how respondents process information in a choice-based conjoint study. They showed that eye tracking is a useful tool to identify whether a certain attribute is considered in a choice task or not. However, the relationship between attention and choice was not totally clear. Furthermore, some visualization techniques that occur often in practice (e.g. pop-ups and hidden information) cannot be investigated using this approach.

This research uses an approach to assess visual impact, in which the effect of an attribute level and its visualization are measured simultaneously in a way that their effects are combined.

1.2 Research questions

The research question of this thesis is: what is the effect of visualization of choice options on the actual choice?

I want to quantify how the visual attribute representations affect consumers’ preferences. The effect of attribute level visualization on consumer choice behavior is explored for three different conditions. For each condition, the hypothesis is described below.

I. Pop-ups and footnotes

Instead of presenting attribute levels using plain and descriptive text, a concise and captivating text is used to cover the most important information. The remaining information is given in a pop-up or footnote.

Hypothesis: using pop-ups and footnotes for a choice option leads to a higher probability that the particular choice option gets chosen.

II. Visible or hidden attributes

Instead of always making attribute levels fully visible, some were hidden, making them only accessible through a “General terms and conditions” window.

Hypothesis: by hiding unimportant information and making it only accessible through a "General terms and conditions" window, this leads to a higher probability that the particular choice option gets chosen.

III. Font size variations

Instead of presenting all attribute levels using the same font size, the font sizes are varied to emphasize or play down attribute levels or aspects of attribute levels.

Hypothesis: using larger font sizes for important attribute levels of a choice option increases the probability that the particular choice option gets chosen, because larger font sizes means more visibility.

Hypothesis: using smaller font sizes for less important attribute levels of a choice option increases the probability that the particular choice option gets chosen, because consumers will mainly disregard less important attribute levels then.

This thesis is of practical relevance, because including visualization results in a research design that better replicates reality, creates more realistic choice exercises and potentially possesses higher external validity. Also, it allows recommendations to be made on how to communicate the level of an attribute to yield the desired results (within ethical boundaries).

1.3 Outline

The outline of this thesis is as follows. In chapter 2, a description of choice-based conjoint analysis is given. This typically results in a multinomial model, which does not take into account differences between respondents. This model is described in chapter 3. The solution to specify the differences is using a mixed logit model. This model is explained in detail in chapter 4. Chapter 5 presents the data for a conjoint study with credit cards. The results of the credit card study are presented in chapter 6. In chapter 7, a discussion on the limitations of this research and some further research directions are given. The conclusions of the whole research project are presented in chapter 8. The questionnaire that is used to gather the data and the script code that was used for the analysis in the open source language R are included in the appendix, to make sure that the reader can understand and reproduce the results.

2. Choice-based conjoint

In this chapter, I describe the basic idea behind the choice-based conjoint method. First, the history of conjoint methods, the choice-based conjoint method in particular, is described in section 2.1. A choice-based conjoint study consists of five consecutive components, starting with the definition of a problem and ending with solving the problem. All these components are covered in section 2.2. In section 2.3, an example of a choice-based conjoint study on the market of 3D televisions is described to give the reader a clear idea of the added value of the method. Finally, the limitations of the choice-based conjoint method can be found in section 2.4.

2.1 The history of choice-based conjoint analysis

The most important characteristic of conjoint methods is that respondents evaluate product concepts, which consist of multiple attributes. Each attribute can have different levels and researchers want to investigate how respondents value the different levels of a product attribute. The objective of conjoint analysis is to determine what combination of levels of the different attributes is most influential on decision making. The history of conjoint analysis starts in 1964, when Luce and Tukey presented their work. Ten years later, this evolved into discrete choice methods by research conducted by McFadden in the field of econometrics.

Paul Green, marketing professor at the University of Pennsylvania, recognized that the study of Luce and Tukey could be applied to the field of market research. It can be used to come to a clear understanding of how consumers make difficult purchase decisions, to estimate the importance of different attribute levels of a product, and it can also be applied to predict consumer purchase behavior. In 1971, Green and Rao published their article 'Conjoint Measurement of Quantifying Judgmental Data', which is based on a full profile card-sort conjoint analysis. This work led to a whole new field in market research. The card-sort approach allows researchers to investigate which attributes were most important and which levels were most preferred. A controlled set of product concepts is shown to respondents in several successive choice tasks, and the valuation of the attribute levels is determined by analyzing the ordering of the product concepts based on attractiveness according to respondents in each choice task. The resulting utilities can be used to develop marketing models that estimate revenues, market shares and profitability of new products. However, this method only works well when the number of attributes is not too long; typically four or five. Researchers came up with different approaches to make it possible to work with product concepts containing more attributes. They came up with methods based on ratings, in which respondents have to answer how much they like each product concept. This can be based on a small number of choice categories ('very much like' to 'very much dislike') to an indication how much they liked it on an interval scale (often from 1-10). One drawback is that the tasks were not realistic and could not be linked directly to purchase behavior (Pullman *et al.*, 1999). As a better alternative to deal with more attributes, hybrid conjoint methods were developed.

Conjoint analysis has gained in popularity among researchers and academics since the early 1980s. This was mainly caused by the possession of computer programming skills. In 1985, commercial software became available. Sawtooth Software, the company of Richard Johnson, released a system called 'adaptive conjoint analysis' (ACA). Johnson discovered that a computer can be programmed to administer a survey and collect data. The computer can adapt the survey to each respondent. In that case, only the most relevant trade-offs are considered, resulting in more realistic responses. Before proceeding to the trade-off phase, respondents have to indicate the importance of the different attributes. The next trade-off questions are based on the previous answers. The goal is to obtain the maximum amount of information regarding the choice behavior of the respondents. A problem with ACA is that the importance of price is underestimated, due to the first part of the interview in which respondents have to state the importance of the different attributes. Often, price seems to be less important when filling out the survey than in reality (Williams and Kilroy, 2000). Various market simulators can be created for what-if-scenarios. Once the preferences of all the respondents for the different levels of the attributes are known, researchers use that to test the acceptance of a product in a competitive environment. In doing so, researchers make use of some strong assumptions. One of these assumptions is called 'additivity'. This means that the value of a product concept equals the sum of the utilities of its attribute levels. This assumption corresponds to a less complex form of decision making, which makes it easier to work with for practitioners.

Shortly before the commercial software programs were released, professor Jordan Louviere of the University of Iowa and his colleagues investigated the usefulness of choice-based approaches to conjoint analysis as a follow-up study of the work of McFadden. They considered making choices among alternatives in combination with the multinomial logit model to solve logistic and marketing problems (Louviere and Woodworth, 1983). This type of analysis seemed to mimic real purchase behavior. Furthermore, it is a better method in case you want to model interactions and cross-elasticity. It resulted in multinomial models of preferences only, because there was not enough information to model preferences for each respondent to incorporate heterogeneity and there were no powerful computers.

The 1990s showed a strong growth for conjoint analyses and its applications. There were many industry usage based studies conducted by academics (Carroll and Green, 1995), which contributed to ACA being the most used conjoint method worldwide. However, choice-based conjoint analysis took over this position. This was mainly caused by the release of new commercial software for discrete choice models by Sawtooth Software in 1993. Another important factor that caused CBC to be more attractive, is the application of hierarchical Bayes (HB) methods to estimate individual-level preferences, which are based on the work of Greg

Allenby of Ohio State University (Allenby and Rossi, 1999). This was beneficial for the accuracy of estimates. The new software made it much easier to conduct CBC studies, while the HB software made the analysis of choice data almost as easy as the analysis of ratings-based conjoint data.

Nowadays, conjoint analysis is applied in various fields; from consumer packaged goods (Desarbo *et al.*, 1995), durables (Wittink *et al.*, 1994), technology products and electronics (Lee *et al.*, 2006) to healthcare (Ryan *et al.*, 1998), and banking services and credit cards (Kara *et al.*, 1994). The research and development in conjoint analysis is focused on the existing methods. Researchers try to reduce the required number of choice tasks for a respondent with more efficient design plans and HB estimation. They try to make the process more realistic by using animated three-dimensional product concepts and by designing virtual shopping environments with realistic shelf spaces (Peral *et al.*, 2012). Furthermore, attempts are being made to fill the gap between choice behavior from interviews and real purchase behavior.

2.2 The process of a choice-based conjoint study

A CBC study consists of five consecutive stages. Figure 2.1 represents an overview of these stages. All stages are briefly discussed in this subsection.

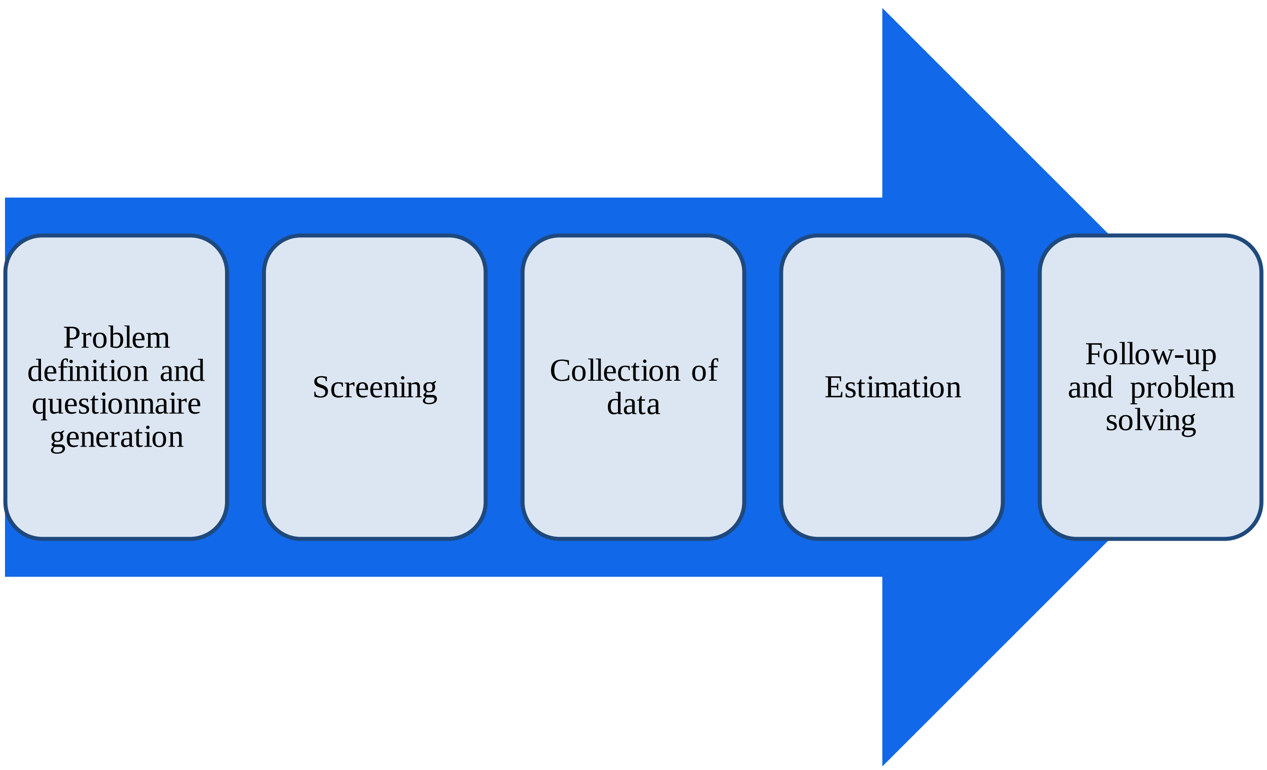


Figure 2.1: An overview of the stages of a CBC study

2.2.1 Problem definition and questionnaire generation

A conjoint study starts with a problem that needs to be solved. Therefore, the characteristics of the market have to be defined and one or more research questions need to be formulated. Typical research questions require measuring price elasticity, finding out which product characteristics individuals consider important, as well as how much they want to pay for those characteristics and products.

First, several product concepts must be created, which consists of several attributes. The number of attributes plays an important role. The consequences of having too few or too many attributes is explained in section 2.4. The respondents should make decisions based on all the attributes. It is likely that the weight of certain attributes is different. For example, if the price of a product is more important than the color of the package, this leads to a higher weight for the attribute "price" compared to the weight for the attribute "color of the package". The recommended number of attributes is four or five at the most (Pullman *et al.*, 1999).

Besides the number of attributes, the number of successive choice tasks and the number of alternatives within each choice task are also important in order to obtain reliable answersThe recommended number of successive choice tasks is thirteen at the most (DeShazo and Fermo, 2002; Iyengar en Lepper, 2000). The motivation for this recommendation is described in section 2.4.

Each attribute consists of a specific number of levels. For conjoint analysis, the levels of a certain attribute must be shown in combination with levels of other attributes. One needs to have information about all the possible relevant combinations of attribute levels. In order to do so, a diverse set of product concepts are created that are shown in specific choice tasks.

2.2.2 Screening

A sample needs to be representative. The characteristics of the respondents in the sample must match the characteristics of the target group. To avoid obtaining biased estimates, the respondents that are being studied have to be chosen carefully. If the respondents have a certain control over whether to participate or not, then this could lead to a self-selection bias. The solution is to use screening. This can be based on various socio-demographic questions, which are useful to include in the questionnaire. Three types of screening can be distinguished: before the respondent answers any of the questions, during the answering of the questions and during the analysis of the data.

The first type of screening is the easiest one to use. It can be useful if a representative sample of a group of people with certain characteristics has to be obtained. Market segmentation can be taken into account. For example, if the preference for the new iPhone has to be determined, 80%

of the respondents must be recent iPhone users and 20% must be users of other mobile phones. This is because iPhone-users are very loyal towards Apple. The other types of screening depend on the respondents' answers to the questions. For example, a researcher may not be very eager to obtain information from people younger than 16 years old, because in that case the parents of the respondents are likely to be responsible for the purchase of the product.

2.2.3 Collection of data

The data needs to be collected from a specific sized sample. This sample size can be small (not more than 100 respondents), but it can also be quite big (at least 1,000 respondents). It may be tempting to determine the required number of respondents, but that makes no sense for the methods that are applied thereafter. Researchers apply a rule of thumb with regard to the statistical sample size and accuracy.

There are two ways of collecting data: completing a paper version of the questionnaire or complete one on the Internet. Nowadays, most of the questionnaires are placed online. The main advantages of placing the questionnaire online is that it is cheaper and less time-consuming. This usually results in more information in a shorter period of time, which is desirable. However, there are a couple of disadvantages as well. These disadvantages and the solutions are explained in detail in section 2.4.

2.2.4 Estimation

After collecting the data, the utilities for the different attribute levels of a product can be estimated. There is more than one model available for this, thus there is no unique way of analyzing the data. The three models that are used most often are:

1. Multinomial logit

2. Latent Class

3. Mixed logit

In this case, the multinomial logit model is chosen; here, the researcher assumes that the satisfaction gained from a level of a product attribute is the same for each respondent. This will result in a single set of utilities for this whole population; aggregate measures are created here. By using an a multinomial logit model, heterogeneity cannot be investigated. That is not very realistic, because no two respondents have the same preferences. Still, this model is a proper basic model to analyze conjoint data. A detailed description of the model is given in chapter 3.

One way to avoid the utilities to be the same for each respondent is by using a latent class model. A class consists of people who are more or less the same. In other words, people within a class share the same preferences. Latent Class estimation detects the different groups of similar respondents (classes) and estimates utilities for each of those. You cannot beforehand determine to which class a person belongs.

Classes can be used for market segmentation. Companies can use different marketing strategies for people belonging to different classes. For example, after collecting the data, a researcher came up with two classes: one class consisted of people who are very sensitive to price changes and the other class consisted of people who consider themselves trendsetters. In that case, a good strategy is to offer the product against a lower price for people from the first class (e.g. by giving them coupons) and by choosing a slightly different design for the product to motivate people from the second class to buy the product (again).

By using classes, heterogeneity is allowed across the classes of respondents only. Hierarchical Bayes applied to choice-based conjoint analysis (CBC HB) is the method that can be used if different utilities for each respondent need to obtained. HB generalizes the Latent Class model. CBC HB can be used to obtain the most information and leads to more accurate and less biased results. The model that is generally used for this method is the mixed logit model. The details of the mixed logit model are described in chapter 4.

2.2.5 Follow up and problem solving

The ultimate goal of the conducted conjoint study is to answer the research questions and to solve the problem. In order to do so, the estimated utilities are needed. These can be used to calculate market shares for different products. An example of a CBC study to calculate market shares is presented in section 2.3. Moreover, you can use certain simulations to get a clear idea of what happens with the market shares of your product and the competitors' market shares when a new (level of an) attribute is added to your consisting product. This allows us for estimating future market shares for new or upgraded products. This is not the only thing that can be done. The researcher could also look at the detected classes and try to come up with good/ better marketing strategies. He can also study interactions between attributes, or look at which attributes are the most important for decision making.

2.3 Example of a choice-based conjoint study

A fictitious example of a CBC study is shown in this subsection to clarify things. In this example, the market of 3D televisions is studied. Product managers can be eager to know how respondents evaluate 3D televisions. Typical research questions are as follows: How important are different levels of the various attributes? How much does a certain level contribute to the probability that a specific product concept is chosen?

First, a choice has to be made between several attributes of a 3D television. As was mentioned before in section 2.2, including too many attributes has a negative impact on the results. The number of attributes must be limited, because otherwise the answers from the respondents are less meaningful. Therefore, only the most important attributes are included. Here, the most important attributes are brand, screen size, PC and internet connection, price, and resolution. Each attribute has different levels. The different levels of the five attributes are displayed in table 2.1. An example of what a product concept would look like is presented in figure 2.2.

Table 2.1: The different attributes and levels of a 3D television.

Attributes	Levels
Brand	LG (B-brand) Philips (A-brand) Samsung (A-brand) Toshiba (B-brand)
Screen size	< 40 inch ≥40 inch
PC + Internet	Yes No
Price	500 euro 1,000 euro 1,500 euro 2,000 euro 2,500 euro 3,000 euro
Resolution	Half HD Full HD

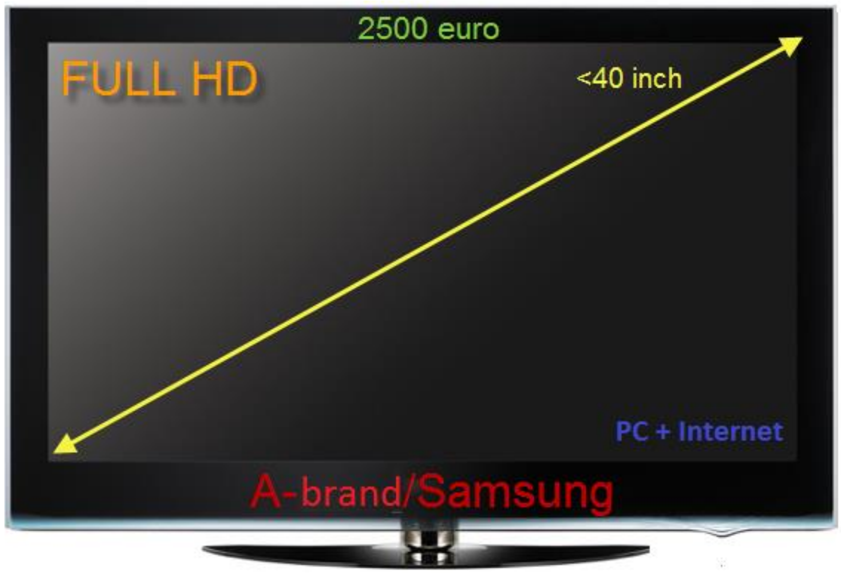


Figure 2.2: The chosen attributes of a 3D television

Consider a single choice task with four alternatives. Each alternative does not need to represent a product which is currently on the market, because CBC allows preferences for new and existing products to be tested. Nevertheless, the resemblance with 3D televisions that are currently on the market is high. A typical choice task is presented in figure 2.3. Respondents are asked to indicate the alternative they want to purchase.

This example shows that there are four completely different product concepts. One can also add a "None"-option. This last option can be useful in case respondents might decline to purchase if neither product concept shown in the choice task is interesting enough for them.

Assume that 100 respondents participated and that they have chosen their most preferred alternative in ten successive choice tasks. These answers are used to perform the estimation. The results of two types of estimation procedures are discussed here; the first is based on the multinomial logit model and the second is the CBC HB estimation procedure. When the multinomial logit model is used, a vector of utilities is obtained. Each level has exactly one utility. When the CBC HB estimation procedure is used, you get a vector of utilities for each respondent. This results into a matrix of utilities. Each row corresponds with the utilities for a specific respondent. To illustrate this, see table 2.2 for the utilities of the levels of the five attributes when the multinomial logit model is used and table 2.3 for the similar results obtained with the CBC HB estimation procedure for the attribute "brand" only. The reason why only the results for this attribute are shown, is to keep things simple and avoid presenting too many utilities (total number of utilities = total number of levels times total number of respondents = 16 x 100 = 1,600).

Choose one of the following

Brand	Samsung
Screen size	< 40 inch
PC + Internet	Yes
Price	1,000 euro
Resolution	Full HD

Brand	Philips
Screen size	≥ 40 inch
PC + Internet	Yes
Price	3,000 euro
Resolution	Full HD

Brand	Toshiba
Screen size	< 40 inch
PC + Internet	No
Price	500 euro
Resolution	Half HD

Brand	LG
Screen size	≥ 40 inch
PC + Internet	Yes
Price	2,000 euro
Resolution	Half HD

Figure 2.3: Example of a CBC choice task

Table 2.2: Utilities obtained with the multinomial logit model

Level of the product attribute	Utility
LG	-0.28
Philips	0.72
Samsung	1.69
Toshiba	-2.13
< 40 inch	-0.42
≥ 40 inch	0.42
Yes	0.23
No	-0.23
500 euro	1.08
1,000 euro	0.92
1,500 euro	0.75
2,000 euro	-0.64
2,500 euro	-0.86
3,000 euro	-1.25
Half HD	-0.53
Full HD	0.53

Table 2.3: Utilities obtained with CBC HB estimation

Level of the attribute 'brand'	LG	Philips	Samsung	Toshiba
Utility respondent 1	-0.08	0.86	2.80	-3.58
Utility respondent 2	-0.24	-0.11	0.62	-0.27
…	…	…	…	…
Utility respondent 100	0.16	-0.28	0.34	-0.22

The average utility of the attribute levels can be calculated for the CBC HB estimation. These values are likely to be about the same as the utilities in table 2.2. The sum of the utilities for all the possible levels of a single attribute, equals 0. This can be seen in table 2.2, because the sum of the first four utilities equals 0. This means that the values are standardized.

Suppose that Samsung wants to know how good its product will sell compared to the products of its competitors. The market shares have to be calculated. First, all the different products available on the market must be defined. To make things easy, we can assume that there are only four different products; these are already described in figure 2.3. It could be stated that the preferences for these products are already known, because the respondents filled in their answers already for the choice task in which they had to make a choice between these four products. For example, if 42 out of the 100 respondents favored the first alternative in this choice task, Samsung's product, the estimated market share equals 42%. The market shares of LG, Philips and Toshiba can be derived the same way. However, these percentages can be wrong. One of the underlying reasons for this is that respondents did not make a "good" choice in this choice task, as it is not consistent with their choices made at other choice tasks. There is more information available based on nine other choice tasks, given by utilities for each level of the product concepts, which leads to more accurate (efficient) estimates of the market shares of the four 3D televisions.

The results are discussed using the multinomial logit model. In order to calculate the market shares of the four different 3D televisions, the total utility of each product must be calculated first. This calculation is based on the 'additivity' property. The total utility equals the sum of the utilities that belong to the levels of the attributes. These values are shown in the first row of table 2.4. To illustrate this, the total utility of product 1 (2.95) is derived by the summation of the utilities that belong to the levels 'Samsung', '< 40 inch', 'Yes', '1,000 euro' and 'Full HD'. By using market shares, the 'share of preference' method is used. First, the exponent of the total utility must be calculated for each product (see the second row of values in table 2.4). Finally, to obtain the market share of a particular product, divide the exponent of the total utility of that product by a normalizing constant. This constant equals the sum of the exponents of the total utilities of all products. So, the market share of a product is proportional to the exponent of the total utility. More information about the calculation steps that are required to end up with the estimated market shares can be found in chapter 3. As you can see, product 1 has the highest market share: 88.5%, followed by product 2 with 8.9%. Product 3 is the least preferred product with only 0.5%.

Table 2.4: Market shares for the multinomial logit model

	Product 1	**Product 2**	**Product 3**	**Product 4**
Total utility	2.95	0.65	-2.23	-0.80
Exponent of total utility	19.106	1.916	0.108	0.450
Market share	88.5%	8.9%	0.5%	2.1%

The resulting market shares for the CBC HB estimation for every single respondent can be derived in a similar way. For each respondent, four values for the total utilities for the different products are calculated. This results into a matrix of 100 rows and four columns with total utilities, see table 2.5. A second matrix, with the same dimensions, can be used to capture the values for the exponents of the total utilities. The share of preferences for the four products for all respondents can be placed in a third matrix, again with the same dimensions.

Table 2.5: Total utilities for the CBC HB estimation

Total utility per respondent	Product 1	Product 2	Product 3	Product 4
Respondent 1	2.68	1.07	-1.55	0.36
Respondent 2	2.89	1.77	-2.56	0.14
…	…	…	…	…
Respondent 100	1.64	0.87	-1.98	-1.02

Samsung may not be particularly interested in the share of preferences, but they want to know the final market shares. An estimation of the final market shares can be obtained with two different methods, which are used to transform the individual part-worth utilities into market predictions:

1. Share of preferences

2. First choice

The share of preferences method assumes that each product has a probability to be chosen. The total probability of all products equals 100%. Again, the share of preferences can be calculated by dividing the exponent of the product's total utility by the normalizing constant. The final market share for a specific product is easily obtained by calculating the mean of the share of preferences for that product from all respondents. This method is especially suitable when studying the market shares for nondurable goods (Desarbo *et al.*, 1995). Consumers are not very loyal to the brand and/ or product; the products that consumers buy vary quite often. One of the reasons is that the taste of the consumers can vary over time, which results in purchasing different products. Share of preferences is preferred in case the purchase of products is not carefully considered or when the buying behavior varies more often. Therefore, in this example of 3D televisions, this method is not used any further.

For the first choice method, we imagine that each respondent is making a choice for the product with the highest total utility. This method makes sense when consumers face more important and less frequent purchases, typically involving more money, such as the purchase of a house or a car; in that case, consumers gather and study relatively more information before they make the decision on which product to buy. The distribution of the preferences for first choice is expected to be flatter than the distribution produced by the method of share of preferences. By comparing the total utilities for the four products in table 2.5, it can be seen that respondent 1 would choose product 3 (3D television of Toshiba), respondent 2 would choose product 1 (3D television of Samsung) and so forth. The final market share of a product is the number of times the product was chosen divided by the number of respondents.

2.4 Limitations of choice-based conjoint

The CBC method is a very useful tool for market analysis. It helps market researchers to analyze all kinds of marketing problems. It provides clear insights in what the market looks like. Forecasts of the future market share can be made, which show the potential power of a product. However, there are some downsides as well. When a researcher chooses to use a CBC method, some caution is advised, especially when the design is decided upon and the results have to be interpreted. In this subsection, the limitations of the CBC methods are described.

2.4.1 Limitations of the respondents

It was already mentioned in section 2.2 that the number of attributes and choice tasks affect the behavior of the respondents. Both too few and too many attributes can have a negative impact on the results. Each product concept should consist of a limited number of attributes that is enough to consider all at once. Too many attributes can lead to respondents considering only a few attributes instead of all the attributes that are shown to them. Too few attributes make the product concept too unrealistic. Consequently, these results are not useful anymore.

Market researchers need to obtain as much information from the respondents as possible. This is required to accurately estimate the utilities for the attribute levels of a product concept. Also, certain combinations of attribute levels must be present in the product concepts. This automatically drives up the required minimal number of choice tasks. Too many choice tasks can lead to a lack of interest and motivation of the respondents. The worst thing that can happen in that case, besides an early termination of the questionnaire, is a random answering of the questions. The results would then no longer be reliable.

The best way to determine whether the respondents' answers are trustworthy is by analyzing the amount of time the respondents took to answer each question. However, caution is important when using this method. The amount of time spent on a choice task decreases when a respondent answered more questions. This can be caused by an increasing lack of interest, but it can also be caused by a gain of experience in answering the questions. Also, respondents that are more motivated need less time to answer, because they are getting more familiar with the exercises and the product concepts. In addition, by rewarding respondents to seriously participate, the reliability increases (Ding *et al.*, 2005). However, this may cause respondents to give answers that are influenced more easily by social desirability or by compromising effects, in which a product concept is chosen in between two others (Rooderkerk *et al.*, 2011).

2.4.2 Interaction effects

In the specification of the CBC method, its attributes should be evaluated independent of each other. For example, in the example study of the 3D televisions, the utilities for each respondent for the resolution “Half HD” or “Full HD” should not depend on the size of the screen. In reality, consumers prefer Full HD more when the screen size is larger; at least 40 inch. This means that there is an interaction between these two attributes. The assumption of independence between two or more attributes is sometimes unrealistic or wrong.

The interaction effect between attributes can easily be estimated. A good approach is to combine these attributes. As for the study of 3D televisions, we have the attributes “resolution” and “screen size”. The corresponding levels are: “Half HD”, “Full HD”, “< 40 inch” and “≥ 40 inch”. The new attribute is the combination of resolution and screen size. It has two times two levels: “Half HD < 40 inch”, “Half HD ≥ 40 inch”, “Full HD < 40 inch” and “Full HD ≥ 40 inch”. You have to be aware of the fact that by allowing for interaction, the total number of levels can increase heavily. To get accurate estimates, the number of product concepts in the CBC must be increased. The effects are presumed to be known now. It is also possible to ignore the interaction between attributes. Market shares are still consistent in this case.

2.4.3 Differences between estimated and actual price sensitivity

One of the advantages of a CBC method, compared to the ACA method, is that it leads to more accurate estimates of the effect of price (Williams and Kilroy, 2000). Market researchers are eager to know the price sensitivity of the respondents, because this can be used to set the optimal price to get the highest profits. Although CBC can be used as a good tool for this purpose, there is a downside here. The estimated price sensitivity in choice tasks can differ from the real price sensitivity in the market. To illustrate this, consider the market of coffee. The majority of the consumers do not really care much about the price of coffee, provided that the price is not too high and they like the taste of the coffee. If a CBC study is conducted for coffee, a respondent is more informed about all the different coffee products. He or she spends more time studying all the features of the coffee and is “forced” to make a consistent decision in each successive choice task. This leads to a higher price sensitivity in the CBC study than in reality. It is not easy to avoid this problem, because the difference in price sensitivity is not observed in practice. The experience with the market can only be used to re-scale the utilities belonging to the attribute price by a certain constant.

2.4.4 Differences between estimated and actual market shares

The predicted market shares in a certain market scenario is never exactly the same as the real market share that can be measured on the market. The goal is to reduce the difference between the two values as much as possible. This difference is caused primarily by external effects and in some degree to a random error. There are many different types of external effects. One of them is promotion. Promotions can make a product more desirable, which can increase the market share. In a CBC study, people have perfect information about all products and only choose a product based on its features. The market share can be underestimated if promotion is taken into account. Therefore, a correction needs to be made.

Another external effect is music in the store. Imagine annoying music being played at the place where the product is sold. This can have a negative impact on the chance to buy the product. Moreover, the place in the store, the shelf space and the space layout even have an effect on the chance to buy the product (Nierop, van, *et al.*, 2008).

Another factor that can be of influence on the difference in market shares is the amount of resources of the respondents. A respondent only indicates what he or she chooses if they have to choose between several products. In reality, a consumer may not even be able to buy the product. It can also happen that the respondent does not want to buy it. A simple solution for this is to add a "None" option to the alternatives in each choice task. Furthermore, the respondent can be someone who is not responsible for the purchases.

Even though there can be a big difference between the estimated market share and the actual market share, the results from the CBC study are still useful, because the market shares are calculated in an idealized condition in which perfect information about all the products is provided. One advantage is that market researchers can now investigate what happens if there are no promotions available. They can also compare the market shares in different scenarios. The differences that result from the comparison lead to information that can be used to undertake strategic decisions.

The choices in a CBC study are dependent of the decision to purchase a particular product from a group of products under consideration. It is difficult to forecast volume sales, especially for new products, because it is not easy to measure the growth of the market. The scenario results of a CBC study are peak results. These results are not reached right away. The product life cycle starts with an introduction to the market and a phase of growth before the maturity phase is reached, in which the sales volume is maximized. The amount of time that is necessary to reach this phase differs among different type of products. On average, it takes less time if the brand of the product is well-known in a positive way and if a new feature of the product is strongly preferred.

In short, CBC methods are very useful for market analysis to study scenarios in which completely new or upgraded products are introduced. The assumption that respondents have perfect information of all the products in a particular product group and that they answer choice tasks as if they buy such a product for the first time leads to valuable information. Although there are some disadvantages, such as neglecting barriers and promotion effects, CBC methods give accurate estimates of market shares and insights in the preferences for certain features of the products that can be exploited more for strategic purposes.

3. The multinomial logit model

In this chapter, I describe a special case of discrete choice models that is used for the choice-based conjoint studies: the multinomial logit model. The probability that a certain alternative within a choice task is chosen can be calculated by using a closed form expression. Utilities are the necessary 'building blocks' to study choice behavior and are thus required for this expression. Therefore, a description of the utilities and their scales is given in section 3.1. The derivation of the closed form expression to calculate the choice probabilities for the multinomial logit model is described in section 3.2. Maximum likelihood (ML) is the estimation procedure that is generally used for the multinomial logit model. The ML estimation procedure is described in section 3.3.

3.1 Discrete choice models

Discrete choice models are models that describe choices made by respondents. Discrete choice models can take on many forms. Examples of prominent discrete choice models are the binary logit model, the multinomial logit model, the nested logit model and the mixed logit model. These models share a number of features, which are described below.

A discrete choice experiment consists of successive choice tasks. A choice task is a set of alternatives that are shown to the respondent. The number of alternatives in each choice task is finite. The respondent chooses exactly one alternative from each choice task. If the researcher uses regression analysis, the dependent variable can take an infinite number of values. This does not hold for discrete choice models. Discrete choice models can be classified according to the number of alternatives in each choice task (assuming that the number of alternatives is the same for every choice tasks). When the number of alternatives in each choice task equals two, one has a binomial choice model. If this number of alternatives is higher than two, it is a multinomial choice model. For example, the number of 3D televisions in the example CBC study in section 2.3 equals four; thus, a multinomial choice model is an appropriate model to use. Another classification can be made for the multinomial choice models. This classification depends on the presence of correlation in unobserved factors among alternatives. In this chapter, the multinomial logit model is used. This model does not allow for correlation among the alternatives.

A discrete choice model describes the choice as depending on some observable characteristics of the alternatives in the choice task and some parameters. These parameters, which are unknown to the market researcher, are called utilities. Utilities describe the satisfaction gained when selecting a particular alternative of a choice task. The researcher is able to determine the utility of a specific level of an attribute, by summing all the utilities for its attribute levels. A decision maker i repeatedly faces a choice among J alternatives. The utility that decision maker i obtains from alternative j is U_{ij}, j = 1, …, J. This utility is unknown to the market researcher. Moreover, the decision maker does not know the exact value either, but he or she can observe which alternative provides the highest value. Under the assumption of utility maximizing behavior, decision maker i chooses alternative j if and only if $U_{ij} > U_{il} \forall l \neq j$.

For the market researcher, it is not possible to observe the utilities of the decision makers, but they can observe the attributes of the alternatives. The attributes of alternative j faced by decision maker i are placed in vector $\boldsymbol{x}_{ij}$. This vector consists of 0/1 values, which represent whether a level of an attribute is absent/present for a specific alternative. For example, for the 3D television study in section 2.3, the (string) vector belonging to the first alternative that consists of five attributes is [‘Samsung’, ‘< 40 inch’, ‘Yes’, ‘1,000 euro’ and ‘Full HD’]’. The levels of this alternative must be recoded into numbers in order to use this information in a econometric model. The total number of attribute levels equals sixteen (see table 2.1). To avoid multicollinearity, the total number of entries for the vector $\boldsymbol{x}_{ij}$ equals the total number of attribute levels minus the total number of attributes. So, the vector $\boldsymbol{x}_{ij}$ has eleven entries.

To illustrate this, it is clarified how this vector is created. It starts with the basic column vector $\boldsymbol{x}_{ij}$, which is a null vector. The first attribute (‘Brand’) has four entries. The brand of the first alternative is ‘Samsung’, which is the third level. So, the third entry gets the value ‘1’. Entries 1 and 2 do not change. A fourth entry for this attribute is not used, because in this case the first three entries are 0, therefore, it is immediately clear that the fourth brand is selected. Thus, three entries are enough for an attribute with four levels. More generally, the number of entries needed for a specific attribute equals the number of levels of this attribute minus one. The second attribute, ‘Screen size’, uses the next entry, because there are only two possible levels for this attribute. The screen size of the first alternative is ‘< 40 inch’. This is the first one of the two levels of the second attribute and so the fourth entry gets the value ‘1’. Following this procedure, the (number) vector $\boldsymbol{x}_{ij}$ corresponding to the first alternative becomes: [0,0,1,1,1,0,1,0,0,0,0]’.

A function can be specified that relates the attributes to the systematic portion of utility (also called ‘representative utility’) of the decision maker V_{ij}, which is one part of the observed utility U_{ij}:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{3.1}$$

$$V_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}, \tag{3.2}$$

where ε_{ij} is a random error term that captures the factors that affect utility U_{ij}, which are not included in V_{ij}. The decomposition of U_{ij} in two parts is fully general, because the random error term ε_{ij} is defined as the difference between the observed utility U_{ij} and the representative utility V_{ij}. In most cases, the market researcher specifies the analytical forms of V_{ij} and ε_{ij} according to the model describing the assumptions about the choices. The vectors $\boldsymbol{x}_{ij}$, for i = 1, …,I and j = 1, …, J, are known, because they describe the choice alternatives. The ([total number of attribute levels – total number of alternatives] x 1) parameter vector $\boldsymbol{\beta}$ is the particularly interesting part for the market researcher. It contains the 'representative utilities for the attribute levels' of the alternative. These values are the primary interest of the research.

The market researcher and the respondents do not know the terms $\varepsilon_{ij}\ \forall\ i, j$, and therefore they are called random error terms. The density $f(\varepsilon_{ij})$ of the random error term is used by the market researcher to make probabilistic statements about the choices of the decision maker. The probability that decision maker i chooses alternative j is:

$$P_{ij} = P(U_{ij} > U_{il}, \forall\ l \neq j), \tag{3.3}$$

$$P_{ij} = P(V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il}, \forall\ l \neq j), \tag{3.4}$$

$$P_{ij} = P(\varepsilon_{il} - \varepsilon_{ij} < V_{ij} - V_{il}, \forall\ l \neq j). \tag{3.5}$$

The probability in Equation (3.5) is a cumulative distribution. If the density $f(\varepsilon_{ij})$ is used, then the cumulative probability can be written as:

$$P_{ij} = \int_{\varepsilon} I(\varepsilon_{il} - \varepsilon_{ij} < V_{ij} - V_{il}, \forall\ l \neq j) f(\varepsilon_{ij})\, d\varepsilon_{ij}, \tag{3.6}$$

where $I(.)$ is an indicator function that equals 1 when the expression in the parentheses is true and 0 otherwise. This is the general expression for the probability for discrete choice models. The density function $f(\varepsilon_{ij})$ is the part of the integral that is different for different discrete choice models. For the CBC method, a multinomial logit model is chosen. This model has closed form expressions for this integral, which make it easier and faster to compute choice probabilities. The derivation of the closed form expression is described in section 3.2.

The probability that a decision maker chooses a particular alternative is determined by comparing the total utilities for all alternatives. The probabilities only depend on the differences in utilities, not on their absolute values. Adding any constant w to the utility of all the alternatives does not lead to a different alternative being chosen:

$$P_{ij} = P(U_{ij} + w > U_{il} + w, \forall\ l \neq j) = P(U_{ij} > U_{il}, \forall\ l \neq j), \tag{3.7}$$

$$P(U_{ij} > U_{il}, \forall\ l \neq j) = P(U_{ij} - U_{il} > 0, \forall\ l \neq j)\,. \tag{3.8}$$

The alternative that had the highest value before remains the alternative with the highest value after adding a constant. Therefore, the absolute values of utilities are meaningless for the market researcher. Since any constant w can be added to the utilities, the absolute values of the utilities can not be estimated. It is possible that different sets of utilities lead to the same choices, for example when the utilities of the first set are always 0.5 higher than the utilities of the second set.

Besides the fact that adding a constant to the utilities does not have an effect on the choice probabilities, multiplying all utilities with a positive constant c does not have an effect on these choice probabilities either. The standard model $U_{ij} = V_{ij} + \varepsilon_{ij}$ is equivalent to the model in which a multiplication factor c is included; $U_{ij}^{c} = cV_{ij} + c\varepsilon_{ij} = c(V_{ij} + \varepsilon_{ij})$ for any positive constant c. The alternative with the highest utility before remains the preferred alternative, no matter how the utilities are scaled.

Utilities have no units and thus it is necessary to normalize the scale of the utilities. This is done by normalizing the variances of the random error terms. When the utilities are multiplied by a positive constant c, the variance of each random error term ε_{ij} is multiplied by c^2; var($c\varepsilon_{ij}$) = c^2var(ε_{ij}). The random error terms are assumed to be independently identically distributed (I.I.D.). This makes it easy to normalize the error variances of all the terms, setting it to a chosen convenient value. When a logit model is used, the error variances are usually normalized to $(\pi^2/6)$. This leads to a small modification of the model that is described by equations (3.1) and (3.2):

$$U_{ij}^{new} = (1/\sigma)\boldsymbol{x}_{ij}'\boldsymbol{\beta} + (1/\sigma)\varepsilon_{ij}, \tag{3.9}$$

where σ = var(ε_{ij}) $/(\pi^2/6)$.

3.2 Derivation of the multinomial logit model

The utility that decision maker i receives from alternative j is presented by U_{ij}, $\forall\ i, j$. This utility can be decomposed in two parts: V_{ij}, which is the systematic portion of utility, and ε_{ij}, which is a random error term that captures the factors that affect utility U_{ij} that are not included in V_{ij}. The multinomial logit model is obtained if each ε_{ij} is independently and identically Gumbel distributed with location parameter μ = 0. The density function and the cumulative distribution function are stated in equations (3.10) and (3.11).

$$f\left(\varepsilon_{ij}\right) = \frac{e^{-\varepsilon_{ij}}}{\sigma} e^{\frac{-\varepsilon_{ij} - \mu}{\sigma}}, \tag{3.10}$$

$$F\left(\varepsilon_{ij}\right) = \exp(-\exp\left(-\frac{\varepsilon_{ij} - \mu}{\sigma}\right)). \tag{3.11}$$

The variance of the distribution is $(\pi^2/6)\sigma$. The scale of the utilities is normalized by dividing the original utilities by σ. The variance of the random error terms are set to the standard values $(\pi^2/6)$. These modifications are described in equation (3.9) in section 3.1.

One of the advantages of using the Gumbel distribution is that the difference between two independently and identically Gumbel distributed variables follows a logistic distribution. If ε_{ij} and ε_{ik} are independently and identically Gumbel distributed, then the difference of the random error terms, ε_{diff} = $\varepsilon_{ij} - \varepsilon_{ik}$, has the following cumulative distribution function:

$$F(\varepsilon_{diff}) = \frac{e^{\varepsilon_{diff}}}{1+e^{\varepsilon_{diff}}}. \tag{3.12}$$

The error terms are independent of each other. This implies that the unobserved portion of utility from one alternative does not provide information about the unobserved portion of utility for another alternative. All the information for the decision process that is used is included in the V_{ij} terms. The remaining portion of the utility can be seen as noise.

The logit choice probabilities can be derived following an indirect approach via Bayes' Theorem (McFadden, 1974). This theorem involves the recovery of the unconditional probabilities from the conditional probabilities. When an event E is considered and $F_1, F_2, \ldots, F_n$ are a collection of discrete events that are mutually exclusive and collectively exhaustive, it can be stated that:

$$P(E) = P(\cup_{i=1}^{n}(E \,\&\, F_i)) = \sum_{i=1}^{n} P(E \,\&\, F_i) = \sum_{i=1}^{n}[P(E|F_i)P(F_i)]. \tag{3.13}$$

In this case, one needs the continuous version. This is the following integral (Jeffreys, 1961):

$$P(E) = \int_{-\infty}^{\infty} f(E|B)f(B)dB,$$

where $f(E|B)$ is the conditional density function of E given B, and $f(B)$ is the marginal density function of B. The density of an event E can be recovered by integration.

Using equation (3.5) from section 3.1, the probability that decision maker i chooses alternative j is defined as:

$$P_{ij} = P(\varepsilon_{il} < \varepsilon_{ij} + V_{ij} - V_{il}, \forall\, l \neq j). \tag{3.14}$$

If the value of the error term ε_{ij} is known, then the expression at the right-hand side of the inequality of equation (3.14) is known. Consequently, the choice probability P_{ij} is conditional on this information. $P_{ij}(\varepsilon_{ij})$ represents the value of P_{ij} given the value of ε_{ij}.

For each l, the cumulative distribution function of ε_{il} evaluated at $\varepsilon_{ij} + V_{ij} - V_{il}$ is given by:

$$F_{\varepsilon_{il}}\left(\varepsilon_{ij} + V_{ij} - V_{il}\right) = \exp(-\exp\left(-\left(\varepsilon_{ij} + V_{ij} - V_{il}\right)\right)). \tag{3.15}$$

All the error terms are independent by definition. This implies that $P_{ij}(\varepsilon_{ij})$ over all $l \neq j$ is the same as the product of the individual cumulative distribution functions, which is given by:

$$P_{ij}\left(\varepsilon_{ij}\right) = \prod_{l \neq j} \exp(-\exp\left(-\left(\varepsilon_{ij} + V_{ij} - V_{il}\right)\right)). \tag{3.16}$$

The ultimate goal is to compute the probability that decision maker i chooses alternative j. This unconditional choice probability can be recovered using Bayes' theorem, which is discussed above. This results in an integral of $P_{ij}\left(\varepsilon_{ij}\right)$ over all values of ε_{ij}, weighted by the density function in equation (3.10):

$$P_{ij} = \int_{-\infty}^{\infty} \left[\prod_{l \neq j} \exp(-\exp\left(-\left(\varepsilon_{ij} + V_{ij} - V_{il}\right)\right))\right] e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}} d\varepsilon_{ij}. \tag{3.17}$$

This integral can be simplified to this compact closed form expression:

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_j e^{V_{ij}}} = \frac{e^{\boldsymbol{\beta}' \boldsymbol{x_{ij}}}}{\sum_j e^{\boldsymbol{\beta}' \boldsymbol{x_{ij}}}}, \tag{3.18}$$

where the term in the numerator involves the representative utility; $V_{ij} = \boldsymbol{\beta}' \boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{ij}$ is a vector of observed variables describing alternative j. The derivation of this closed form expression is presented below and starts with equation (3.17):

$$P_{ij} = \int_{-\infty}^{\infty} \left[\prod_{l \neq j} \exp(-\exp\left(-\left(\varepsilon_{ij} + V_{ij} - V_{il}\right)\right))\right] e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}} d\varepsilon_{ij}.$$

First, the restriction $l \neq j$ in the product term must be eliminated. When $l = j$, the terms V_{ij} and V_{il} are the same. Adding the j term to the product and then multiplying it with $e^{e^{-\varepsilon_{ij}}}$yields

$$P_{ij} = \int_{-\infty}^{\infty} \left[\prod_{l} \exp(-\exp\left(-\left(\varepsilon_{ij} + V_{ij} - V_{il}\right)\right))\right] e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}} e^{e^{-\varepsilon_{ij}}} d\varepsilon_{ij}. \tag{3.19}$$

The two terms at the end cancel out, which leaves the reduced form

$$P_{ij} = \int_{-\infty}^{\infty} \left[\prod_{l} \exp(-\exp\left(-\left(\varepsilon_{ij} + V_{ij} - V_{il}\right)\right))\right] e^{-\varepsilon_{ij}} d\varepsilon_{ij}. \tag{3.20}$$

The product within the square brackets can be simplified as follows:

$$\prod_l \exp(-\exp\left(-\left(\varepsilon_{ij} + V_{ij} - V_{il}\right)\right)) = \exp\left[-\sum_l e^{-(\varepsilon_{ij} + V_{ij} - V_{il})}\right] \tag{3.21}$$

$$= \exp\left[-e^{-\varepsilon_{ij}} \sum_l e^{-(V_{ij} - V_{il})}\right] = \exp[-e^{-\varepsilon_{ij}} W],$$

where $W = \sum_l e^{-(V_{ij} - V_{il})}$, which is independent of ε_{ij}.

Then, the choice probability can be written as:

$$P_{ij} = \int_{-\infty}^{\infty} \exp[-e^{-\varepsilon_{ij}} W]\, e^{-\varepsilon_{ij}} d\varepsilon_{ij}. \tag{3.22}$$

The computation of this integral can be simplified using the Jacobian method. The first step involves the transformation $a = e^{-\varepsilon_{ij}}$. Consequently, the new lower bound of the integral is 0 and the new upper bound remains ∞. The inverse transformation results in $\varepsilon_{ij} = -\ln a$. The Jacobian J is obtained by the division of $d\varepsilon_{ij}$ by da and this results in $-1/a$. The value of a is positive by definition, because the exponent of a number is always positive. The absolute value of the Jacobian is $|J| = 1/a$. This leads to:

$$P_{ij} = \int_0^{\infty} e^{-Wa} a |J|\, da = \int_0^{\infty} e^{-Wa}\, da. \tag{3.23}$$

The computation of this integral is trivial. The closed form expression in equation (3.18) is obtained by substituting the right definition of W, and by means of simplification.

$$P_{ij} = {}_0^{\infty}\left|\left(\frac{-e^{-Wa}}{W}\right) = \frac{1}{W} = \frac{1}{\sum_l e^{-(V_{ij} - V_{il})}}\right. \tag{3.24}$$

$$= \frac{1}{e^{-V_{ij}} \sum_l e^{V_{il}}} = \frac{e^{V_{ij}}}{\sum_j e^{V_{ij}}} = \frac{e^{\boldsymbol{\beta}' \mathbf{x_{ij}}}}{\sum_j e^{\boldsymbol{\beta}' \mathbf{x_{ij}}}}.$$

The simple closed form expression is one of the biggest advantages of using a logit model over other discrete choice models. For example, if a market researcher chooses a probit model, it becomes more difficult to calculate the choice probabilities. That can be explained by the fact that the probabilities in a probit model with J alternatives have to be approximated by the value of a J-tuple integral.

Using a multinomial logit model provides a fast and easy way to calculate the choice probabilities. This is especially important when simulation-based estimations have to be performed. The estimation procedure for the multinomial logit model is described in section 3.3.

3.3 Estimation procedure of the multinomial logit model

The choice probabilities are estimated using the maximum likelihood estimation procedure. This leads to the maximum likelihood estimate, which is the optimal estimate for the choice-based conjoint studies. However, it can be biased (Firth, 2008).

A sample of I decision makers is used, which is randomly selected from the population. There are K successive choice tasks and each choice task has J alternatives (the number of alternatives is the same for each choice task). Every decision maker must choose one alternative in each choice task.

The probability that decision maker i chooses the alternative j that was actually observed as the chosen alternative in the single choice task k is given by the following formula:

$$\prod_{l=1}^{J} P_{ilk}{}^{y_{ilk}}, \tag{3.25}$$

where y_{ijk} = 1 if decision maker i chooses alternative j in choice task k and 0 otherwise.

The choices in each choice task are independent of each other. The probability that the actual choices in K different choice tasks is observed is:

$$\prod_{k=1}^{K} \prod_{l=1}^{J} P_{ilk}{}^{y_{ilk}} \tag{3.26}$$

$$= \prod_{m=1}^{M} P_{im}{}^{y_{im}},$$

where P_{im}, m = 1, …, $M = K * J$, is the total set of choice probabilities of each alternative in each choice task for decision maker i and y_{im} = 1 if decision maker i chooses alternative m and 0 otherwise.

The estimation procedure is not based on one decision maker only, but on the total selected sample. By assuming that the choices of each decision maker are independent of the choices made by other decision makers, the probability of each decision maker in the sample producing the observed choices is given by the likelihood function in the following equation:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{I} \prod_{m=1}^{M} P_{im}{}^{y_{im}}, \tag{3.27}$$

where $\boldsymbol{\beta}$ is the vector that contains the parameters of the model. This likelihood function has a unique maximum. Market researchers often use the log-likelihood function, because it leads to a faster optimization compared to the likelihood function. The reason for this is that the log-likelihood function is globally concave for linear parameters (McFadden, 1974).

The log-likelihood function is obtained by taking the logarithm of the expression in equation (3.27):

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \ln P_{im}. \tag{3.28}$$

The maximum likelihood estimate (MLE) is the value of $\boldsymbol{\beta}$, which maximizes this function. The MLE is obtained by setting the first derivative of the log-likelihood function to 0 and solving for $\boldsymbol{\beta}$. This first order condition (F.O.C.) is given by the following equation:

$$\frac{dl(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = 0. \tag{3.29}$$

Using the closed-form expression for the logit model for the choice probabilities, the F.O.C. can be written in such a way that the expression is easy to interpret. This research starts with equation (3.28) and rewrites it to end up with an expression, which is used to take the derivative from.

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \ln P_{im} \\ &= \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \ln \frac{e^{\boldsymbol{\beta}' \boldsymbol{x}_{im}}}{\sum_j e^{\boldsymbol{\beta}' \boldsymbol{x}_{ij}}} \\ &= \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} (\boldsymbol{\beta}' \boldsymbol{x}_{im}) - \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \ln \sum_j e^{\boldsymbol{\beta}' x_{ij}}. \end{aligned} \tag{3.30}$$

The derivative of this log-likelihood function is:

$$\begin{aligned} \frac{dl(\boldsymbol{\beta})}{d\boldsymbol{\beta}} &= \frac{d(\sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} (\boldsymbol{\beta}' \boldsymbol{x}_{im}) - \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \ln \sum_j e^{\boldsymbol{\beta}' x_{ij}})}{d\boldsymbol{\beta}} \\ &= \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \boldsymbol{x}_{im} - \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \sum_j P_{ij} \boldsymbol{x}_{ij} \\ &= \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \boldsymbol{x}_{im} - \sum_{i=1}^{I} (\sum_j P_{ij} \boldsymbol{x}_{ij}) \sum_{m=1}^{M} y_{im} \\ &= \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \boldsymbol{x}_{im} - \sum_{i=1}^{I} (\sum_j P_{ij} \boldsymbol{x}_{ij}) \\ &= \sum_{i=1}^{I} \sum_{m=1}^{M} (y_{im} - P_{im}) \boldsymbol{x}_{im} \\ &= \sum_{i=1}^{I} \sum_{m=1}^{M} y_{im} \boldsymbol{x}_{im} - \sum_{i=1}^{I} \sum_{m=1}^{M} P_{im} \boldsymbol{x}_{im}. \end{aligned} \tag{3.31}$$

The first order condition is obtained by setting this derivative to 0. This implies that the two terms of the derivative must be the same in the optimum:

$$\sum_{i=1}^{I}\sum_{m=1}^{M} y_{im}\boldsymbol{x}_{im} = \sum_{i=1}^{I}\sum_{m=1}^{M} P_{im}\,\boldsymbol{x}_{im} \tag{3.32}$$

Dividing both terms by I leads to the following expression:

$$\frac{1}{I}\sum_{i=1}^{I}\sum_{m=1}^{M} y_{im}\boldsymbol{x}_{im} = \frac{1}{I}\sum_{i=1}^{I}\sum_{m=1}^{M} P_{im}\,\boldsymbol{x}_{im}. \tag{3.33}$$

The left-hand side of the equation represents the average of $\boldsymbol{x}$ over the alternatives chosen by the sampled decision makers; the average of $\boldsymbol{x}$ over the predicted choices of the sampled decision makers is represented at the right hand. The maximum likelihood estimate of $\boldsymbol{\beta}$ is obtained in such a way that that the predicted average of every explanatory variable is equal to the observed average in the sample. The estimates are obtained in such a way that the model reproduces the observed averages in the sample. In most cases, the explanatory variables can be recoded to dummy variables. This implies that the left-hand side of equation (3.33) presents the share of people in the sample who have chosen alternative j. The predicted share of people who would have chosen alternative j is given at the right-side of the equation and must be exactly the same.

The multinomial logit model results in estimates of $\boldsymbol{\beta}$, which are the same for each decision maker. However, there is a great variety in tastes of people and it is interesting to know the individual preferences rather than the preferences of the average people. Therefore, a mixed logit model is used that incorporates heterogeneity. The estimation procedure for this model is the hierarchical Bayes (HB) procedure for choice-based conjoint analysis. The mixed logit model and the HB estimation procedure are described in detail in chapter 4.

4. The mixed logit model

The multinomial logit model does not allow for respondent heterogeneity. Every respondent is treated the same way, because the multinomial logit model is only concerned about the average preferences of people. There is a great variety in tastes of people and it is particularly interesting for a market researcher to know the individual preferences (Sandor and Wedel, 2005). This allows the market researcher to discover market segments that can be targeted strategically to reach the goals of a company. The mixed logit model is used to incorporate heterogeneity. It allows respondents to have their own tastes. The model assumes that each respondent is in fact a random sample from an underlying population. The distribution of this population is an important feature of the mixed logit model.

In section 4.1, the mixed logit model is described in detail; this is the central model in this thesis. The mixed logit model can be estimated using Bayesian procedures. In section 4.2, an overview of Bayesian concepts and properties is presented. In section 4.3, the hierarchical Bayes procedure for estimating the mixed logit model is described in detail. Choice-based conjoint studies frequently include product attributes for which almost everybody would be expected to prefer a specific level to another. However, the estimated individual part-worth utilities could have a different order. This can lead to several problems; however, these can be avoided by using constraints. The possibility to and consequences of implementing constraints for the utilities of attribute levels of an attribute are explored in section 4.4.

4.1 The mixed logit model

The mixed logit model is based on choice probabilities at the individual level. Several definitions and notations from chapter 3 are readdressed in this section. The main difference between the multinomial logit model and the mixed logit model is that, rather than obtaining a vector of parameters $\boldsymbol{\beta}$ for the whole population, each respondent i has his own vector of parameters $\boldsymbol{\beta}_i$.

The mixed logit model for the discrete choice conjoint estimation is a hierarchical model existing of two levels: an upper level and a lower level. The upper level assumes that the respondents' utilities are distributed by a multivariate normal distribution. The lower level assumes that the individual choices are described by a logit model. The upper level represents the sample drawn from the population and the lower level represents a sample of choice task evaluations for a respondent. The utility that respondent i obtains from alternative j in choice task k is given by:

$$U_{ijk} = \boldsymbol{\beta}_i' \boldsymbol{x}_{ijk} + \varepsilon_{ijk}, \qquad (4.1)$$

where $\boldsymbol{x}_{ijk}$ is the (dummy) vector for the attribute levels of alternative j in choice task k for respondent i, ε_{ijk} is I.I.D. extreme value, and $\boldsymbol{\beta}_i \sim N(\boldsymbol{b}, \boldsymbol{W})$ with a vector of means $\boldsymbol{b}$ and a covariance matrix $\boldsymbol{W}$.

A sample of I respondents is observed. Suppose that the total number of choice tasks (K) is the same for each respondent. The chosen alternatives for all the choice tasks for decision maker i are denoted by $\boldsymbol{y}_i' = [y_{i1}, y_{i2}, \dots, y_{iK}]$. The choices of the entire sample are placed in the matrix $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_I]$.

The probability of respondent i's observed choices, conditional on the individual parameter vector $\boldsymbol{\beta}_i$, is:

$$L(\boldsymbol{y}_i \mid \boldsymbol{\beta}_i) = \prod_{k=1}^{K} \left(\frac{e^{\boldsymbol{\beta}_i' x_{i y_{ik} k}}}{\sum_{l=1}^{J} e^{\boldsymbol{\beta}_i' x_{ilk}}} \right), \tag{4.2}$$

where $y_{ik} = j$ if, and only if, respondent i chose alternative j in choice task k.

The probability that is not conditional on the individual parameter vector $\boldsymbol{\beta}_i$ is the integral of $L(\boldsymbol{y}_i \mid \boldsymbol{\beta}_i)$ over all values of $\boldsymbol{\beta}_i$:

$$L(\boldsymbol{y}_i \mid \boldsymbol{b}, \boldsymbol{W}) = \int L(\boldsymbol{y}_i \mid \boldsymbol{\beta}_i) \phi(\boldsymbol{\beta}_i \mid \boldsymbol{b}, \boldsymbol{W}) \, \mathrm{d}\boldsymbol{\beta}_i, \tag{4.3}$$

where $\phi(\boldsymbol{\beta}_i \mid \boldsymbol{b}, \boldsymbol{W})$ is a normal density with mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{W}$. The term $L(\boldsymbol{y}_i \mid \boldsymbol{b}, \boldsymbol{W})$ is called the mixed logit probability, which lends its name to the model.

One of the estimation procedures for the mixed logit model is a hierarchical Bayesian procedure. Section 4.2 describes the Bayesian concepts and properties.

4.2 Bayesian procedures and properties

Bayesian procedures can be used to overcome some difficulties associated with the classical procedures. One of the reasons of the popularity of Bayesian procedures is that it is not required to maximize any function (Zellner, 1971). In contrast, maximum likelihood requires to maximize numerically. For complex cases, there is a probability that there is no convergence at all and algorithms converge to different (local) maxima for different starting values. Bayesian procedures can also have problems with convergence, but in another setting; they require an iterative process for estimating values that can converge only after a sufficient number of iterations.

Another advantage of Bayesian procedures is that the properties ‘consistency’ and ‘efficiency’ are more easily obtained than with classical procedures. For Bayesian estimators, consistency is obtained using only a fixed number of draws; the estimators are efficient if the number of draws rises at any rate with the total sample size (Bauwens *et al.*, 1999). More information about draws is given at the end of this paragraph. In case of limited data sets, asymptotic theory is not valid, but Bayesian inference is exact even in small samples. Also, Bayesian procedures lead to increases in forecasting power compared to the classical approaches (Rossi *et al.*, 2006).

The relative speed of both procedures is important for the decision on which procedure to use. Increases in computing power makes Bayesian analysis operational. Also, Bayesian procedures are usually easier to program.

It is possible to look at statistical inference in another way using the Bayesian approach, as it is completely different from the classical (frequentist) approach. Consider a one-dimensional model with parameter θ and let $\hat{\theta}$ be an estimate of θ and $\hat{s}_{\hat{\theta}}$ its estimated standard error. If the classical approach is used, there is a 95% probability that the interval $(\hat{\theta} - 1.96\hat{s}_{\hat{\theta}},\ \hat{\theta} + 1.96\hat{s}_{\hat{\theta}})$ contains the true parameter value θ_0. The Bayesian interpretation is as follows: the probability that the true parameter value θ_0 lies in the interval $(\hat{\theta} - 1.96\hat{s}_{\hat{\theta}},\ \hat{\theta} + 1.96\hat{s}_{\hat{\theta}})$ is 95%. In the Bayesian approach, the parameter is treated as a random variable. The probability describes the state of knowledge about the true parameter value θ_0.

4.2.1 Prior distribution, likelihood function and posterior distribution

The market researcher can have some initial ideas about the parameter value prior to collecting data. This is most often based on economic theory, intuition, past analyses and experts’ opinions. The ideas about the parameter are represented by a probability distribution concerning all the possible values for the parameter. The probability corresponding to a specific value measures how likely the market researcher thinks it is for the parameter to take that value. The probability distribution is called the ‘prior distribution’ and is denoted by $\pi(\theta)$. The market researcher collects data, which is used to improve the information contained within the prior distribution. A sample of I independent decision makers is observed; this results in the observed choices y_i, for i = 1, … , I. The set of observed choices in the total sample is denoted by $\boldsymbol{Y} = \{y_1, y_2, \dots, y_I)$. This set is used to update the researcher’s ideas about the parameters. The updated version of the ideas about the parameters is presented by a new density function $\pi(\theta|Y)$, which is called the ‘posterior distribution’.

The relationship between the prior distribution and the posterior distribution is established by Bayes’ rule. Let $P(y_i|\theta)$ be the probability of observing outcome y_i for decision maker i. This probability can be expressed in terms of explanatory variables, but that is omitted for sake of simplicity. By assuming independence, the probability of observing the set of outcomes Y is the likelihood function:

$$L(\boldsymbol{Y}|\theta) = \prod_{i=1}^{I} P(\,y_i|\theta). \tag{4.4}$$

The posterior distribution is related to this likelihood function, which can be shown by using the rules of conditioning:

$$\pi(\theta|\boldsymbol{Y})f(\boldsymbol{Y}) = L(\boldsymbol{Y}|\theta)\pi(\theta), \tag{4.5}$$

where the function $f(\boldsymbol{Y})$ is the marginal probability of $\boldsymbol{Y}$.

Equation (4.5) shows that the posterior distribution $\pi(\theta|\boldsymbol{Y})$ can be expressed in terms of the prior distribution $\pi(\theta)$, the likelihood function $L(\boldsymbol{Y}|\theta)$ and the marginal density function $f(\boldsymbol{Y})$. A small rearrangement leads to the following expression for the posterior distribution:

$$\pi(\theta|\boldsymbol{Y}) = \frac{L(\boldsymbol{Y}|\theta)\pi(\theta)}{f(\boldsymbol{Y})}. \tag{4.6}$$

This equation can be made more compact, because the denominator is simply a normalizing constant with respect to the parameter θ that assures that the posterior distribution integrates to 1. Therefore, it can be said that the posterior distribution is proportional to the likelihood function times the prior distribution:

$$\pi(\theta|\boldsymbol{Y}) \propto L(\boldsymbol{Y}|\theta)\pi(\theta). \tag{4.7}$$

This implies that the probability a market researcher ascribed to a certain value for the parameter θ after seeing the total sample, is proportional to the probability that is ascribed before seeing the total sample times the probability that this parameter value would have resulted in the observed choices.

One way to report the results is by making a graph of the prior and posterior densities of θ. In figure 4.1, a fictitious example of three graphs for the prior distribution, the likelihood function, and the posterior distribution are displayed. It is shown that the posterior distribution is almost similar to the likelihood function. This implies that the likelihood function dominates here and the effect of the prior is small. The effect of the likelihood function is expected to increase when more data becomes available, which leads to a more peaked posterior distribution with almost the same mean as the likelihood function.

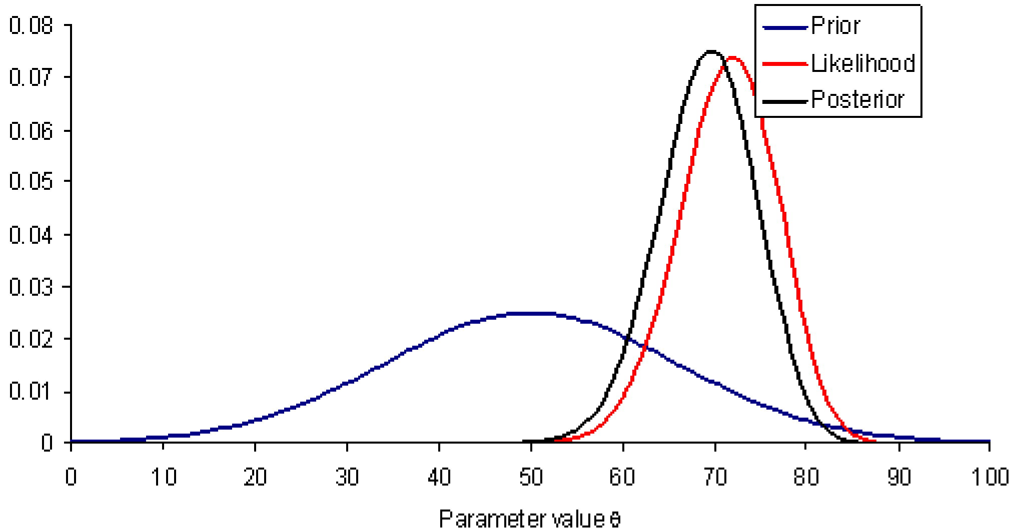


Figure 4.1: Graphs for the prior distribution, likelihood function and the posterior distribution

4.2.2 The mean of the posterior distribution

Graphs are useful for reporting when the number of parameters is small. This is not feasible with many parameters, as the results should be presented in a clear and concise way. A solution for this is to report point estimates of θ instead of the whole distribution. The most common point estimate is the mean of the posterior distribution:

$$\bar{\theta} = \int \theta \pi(\theta|\boldsymbol{Y}) d\theta. \tag{4.8}$$

From a Bayesian perspective, the posterior mean is the value of θ that results from the minimization of the expected cost of the market researcher being wrong about θ, if the cost of error is a quadratic function of the size of the error. Therefore, the cost of being wrong about the true value of θ rises quadratically with the distance to the true value of θ. A market researcher can also choose other loss functions, such as the absolute loss function or the zero-one loss function. The corresponding point estimates are the median and the mode of the posterior distribution (Zellner, 1971). It could even be decided to use loss functions that are bounded and asymmetric (Wen and Levy, 2001). The choice for a specific loss function depends on the posterior distribution. In this thesis, the quadratic loss function and thus the posterior mean are used, because the costs of the errors should rise if the absolute size of the error increases.

The classical approach is concerned with the determination of the sampling distribution of an estimator. Different samples produce different point estimates and the sampling distribution is the distribution of point estimates. This distribution is obtained when enough different samples were taken. The asymptotic sampling distribution is an approximation of the actual sampling distribution when the sample size is large enough. Market researchers that use the classical approach want to find out what the sampling distribution of (the statistic) $\bar{\theta}$ looks like. The answer can be given using the Bernstein – von Mises – theorem. This theorem, which is named after Bernstein (1917) and von Mises (1931), provides the proof of the observation of Laplace (1820): the posterior distribution resembles the normal distribution as the sample size increases. An interesting finding of this theorem is that it is established that the posterior mean is similar to the maximum likelihood estimator. Instead of trying to maximize the likelihood function, a market researcher can also calculate the mean of the posterior distribution to end up with an estimator that is as good (in classical terms) as maximum likelihood. However, there could be a difference between the posterior mean and the maximum of the likelihood function when the sample size is insufficient for asymptotic convergence (Sonnier *et al.*, 2007). When there is such a difference, other reasons to motivate the decision on which one to choose (if a choice is necessary) should be searched, because they both have the same asymptotic properties.

4.2.3 Taking draws from the posterior distribution

The posterior mean can be calculated according to the formula in equation (4.8). An approximation can be obtained by taking A draws of θ from the posterior distribution $\pi(\theta|Y)$ and averaging the results. The simulated mean of the posterior is:

$$\tilde{\theta} = \frac{1}{A}\sum_{a=1}^{A}\theta^{(a)}, \tag{4.9}$$

where $\theta^{(a)}$ is the a-th draw from the posterior distribution $\pi(\theta|Y)$. The simulated mean of the posterior $\tilde{\theta}$ is consistent and asymptotically normal distributed for a fixed A and becomes efficient and equivalent to maximum likelihood if A rises at any rate with the sample size. This only holds when the draws from the posterior distribution are independent and when they can be taken without having to simulate the choice probabilities.

Taking draws is easy for a one-dimensional distribution, but in most cases there is more than one parameter. In case of a high-dimensional distribution, the parameters are placed in a parameter vector. It is rare that the posterior takes on the same form for each of the parameters within the entire parameter vector. There are several ways to deal with this type of complex problems, which are computationally faster and suited better than standard sampling methods. Two popular methods that are especially useful for taking draws from a posterior distribution are “Gibbs sampling” and the “Metropolis – Hastings method”. These methods are called “Monte Carlo Markov chain” (MCMC) methods, because they are based on Markov theory. A Markov chain can be constructed when the limiting distribution equals the posterior distribution. An iterative procedure is used, which means that the next parameter value is dependent of the previous one. After convergence, draws can be used to calculate the posterior mean. In section 4.3, a way to check whether the MCMC has converged or not is presented.

4.2.4 Gibbs sampling

Gibbs sampling does not require the user to take draws from the multidimensional posterior distribution for all parameters at the same time. It allows to take draws of one parameter at a time, conditional on the values on the other parameters (Casella and George, 1992). The idea of Gibbs sampling is to draw iteratively from the full conditional posterior distributions. This is much easier than simultaneously drawing from the posterior distributions for all parameters. Gibbs sampling for the parameter vector $\boldsymbol{\theta}$ is done as follows:

1. Starting values $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots,\ \theta_M^{(0)})$ and set q = 0.

2. Simulate
$\theta_1^{(q+1)}$ from $\pi(\theta_1|\theta_2^{(q)}, \theta_3^{(q)}, \ldots, \theta_M^{(q)}, \boldsymbol{Y})$
$\theta_2^{(q+1)}$ from $\pi(\theta_2|\theta_1^{(q+1)}, \theta_3^{(q)}, \ldots, \theta_M^{(q)}, \boldsymbol{Y})$
$\theta_3^{(q+1)}$ from $\pi(\theta_3|\theta_1^{(q+1)}, \theta_2^{(q+1)}, \theta_4^{(q)}, \ldots, \theta_M^{(q)}, \boldsymbol{Y})$
$\vdots$
$\theta_M^{(q+1)}$ from $\pi(\theta_M|\theta_1^{(q+1)}, \theta_2^{(q+1)}, \ldots, \theta_{M-1}^{(q+1)}, \boldsymbol{Y})$.

3. Set q = q + 1 and go to step 2.

After convergence (q = q^*), the simulated values $\{\boldsymbol{\theta}^{(q)},\ q\ \geq\ q^*\}$ are used as a sample from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ (Tierney, 1994).

4.2.5 Metropolis – Hastings algorithm

The Metropolis – Hastings sampler is particularly useful when the full conditional posterior distributions are unknown. In fact, Gibbs sampling is just one type of a Metropolis – Hastings algorithm (Gelman and Rubin, 1992). To keep things simple, the term Metropolis – Hasting is only used here for versions that are more complex than Gibbs sampling. The Metropolis – Hastings algorithm is especially useful for posterior distributions, because it does not require the user to calculate the normalizing constant in equation (4.6) which is sometimes hard to calculate.

The Metropolis – Hastings algorithm uses the candidate-generating density function $g\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(q)}\right)$. Different choices for the candidate-generating function results in different specific forms of the algorithm. The steps of the (random walk) Metropolis – Hastings sampler for draws of the posterior distribution are as follows:

1. Starting values $\boldsymbol{\theta}^{(0)}$ are specified and $q = 0$.

2. Simulate $\boldsymbol{\theta}^*$ from $g\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(q)}\right)$.
   Set $\boldsymbol{\theta}^{(q+1)} = \boldsymbol{\theta}^*$ with probability α,
   and $\boldsymbol{\theta}^{(q+1)} = \boldsymbol{\theta}^{(q)}$ with probability $1 - \alpha$,
   where $\alpha = \min\left(\frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{Y})g(\boldsymbol{\theta}^{(q)}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{(q)}|\boldsymbol{Y})g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(q)})}, 1\right)$.

3. Set $q = q + 1$ and go to step 2.

The acceptance probability α gives the fraction of candidates that are accepted. As a rule of thumb, the acceptance probability is between 50% (one-dimensional sampler) and 25% (high-dimensional sampler) (Lancaster, 2004). After convergence ($q = q^*$), the simulated values $\{\boldsymbol{\theta}^{(q)}, q \geq q^*\}$ are used as a sample from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ (Chib and Greenberg, 1995).

In the following section, the Bayesian estimation procedure for the mixed logit model is described. Applications of the Gibbs sampling and Metropolis – Hastings algorithm are used.

4.3 Estimation procedure for the mixed logit model

In this section, the Bayesian procedures that are used to estimate the parameters of the mixed logit model are presented. The approach that is mainly developed by Allenby (1995) is utilized. This approach has been generalized afterwards by Train (2001).

4.3.1. The posterior distribution for the mixed logit model

The normal distribution of $\boldsymbol{\beta}_i$ allows for a fast estimation, because "conjugate" priors are used. This means that the distribution of the prior has the same form as the distribution of the posterior. Each respondent is a draw from the population and the respondents are exchangeable. When a market researcher wants to make a prediction for a new respondent, he can regard this new respondent as a draw from the same population. The vectors $\boldsymbol{\beta}_i$, for $i = 1, \dots, I$, are I.I.D. draws from the normal distribution. In short, the respondents are a random sample of the population whose utilities are normally distributed with mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{W}$.

If the market researcher wants to apply a Bayesian estimation procedure, he must give priors for mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{W}$. The prior on $\boldsymbol{b}$ is chosen to be normal. If there is no prior information available on $\boldsymbol{b}$, an unbounded large variance can be chosen, such that the prior on $\boldsymbol{b}$ is a non-informative prior. An inverted Wishart distribution is chosen as the prior for $\boldsymbol{W}$. In case of no prior information, the shape parameters $\boldsymbol{M}$ and the scale matrix $\boldsymbol{Q}_M$ are chosen, where

$\boldsymbol{M}$ is the dimension of the parameter space and $\boldsymbol{Q}_M$ the $\boldsymbol{M}$ – dimensional scale matrix. One may also choose a more flexible prior specification for $\boldsymbol{W}$, but that makes the Gibbs sampling more complex (McCulloch *et al.*, 2000).

The posterior distribution is proportional to the likelihood function times the prior distribution. The likelihood function is obtained by calculating the product of all the individual likelihood functions. The prior distribution, $\pi(\boldsymbol{b},\boldsymbol{W})$, is obtained by the product of the prior distribution of $\boldsymbol{b}$ and the prior distribution of $\boldsymbol{W}$:

$$\pi(\boldsymbol{b},\boldsymbol{W}) = \pi(\boldsymbol{b}) * \pi(\boldsymbol{W}), \tag{4.10}$$

where $\pi(\boldsymbol{b})$ is a normal distribution for $\boldsymbol{b}$ and $\pi(\boldsymbol{W})$ is an inverted Wishart distribution for $\boldsymbol{W}$.

The posterior distribution of $\boldsymbol{b}$ and $\boldsymbol{W}$ is:

$$\pi(\boldsymbol{b},\boldsymbol{W}\,|\boldsymbol{Y}) \propto \prod_{i=1}^{I} L(\boldsymbol{y}_i \mid \boldsymbol{b},\boldsymbol{W})\ \pi(\boldsymbol{b},\boldsymbol{W}). \tag{4.11}$$

Taking draws directly from this posterior distribution is possible using the Metropolis – Hastings algorithm, but this is very time-consuming. This algorithm requires calculating the right-hand side of equation (4.11) for each iteration. The mixed logit probability is an integral without a closed form, which must be approximated using simulation. This implies that simulation of the mixed logit probability is needed for each respondent i for every iteration. Furthermore, the properties of the resulting estimator are affected by this approach, because the properties of the simulated posterior mean are derived assuming that draws are taken from the posterior distribution without the need of simulating choice probabilities. The solution for a computationally faster way of taking draws is to consider all $\boldsymbol{\beta}_i$s as parameters along with $\boldsymbol{b}$ and $\boldsymbol{W}$. Consequently, the posterior distribution is:

$$\pi(\boldsymbol{b},\boldsymbol{W},\boldsymbol{\beta}_i\ \forall\ i\,|\boldsymbol{Y}) \propto \prod_{i=1}^{I} L(\boldsymbol{y}_i \mid \boldsymbol{\beta}_i)\, \phi(\boldsymbol{\beta}_i |\, \boldsymbol{b},\boldsymbol{W})\pi(\boldsymbol{b},\boldsymbol{W}). \tag{4.12}$$

Draws from this posterior can be obtained using a Metropolis – Hastings algorithm. The draws of each parameter are taken conditional on the other parameters. This is done in three steps:

1. $\boldsymbol{b} \mid W, \boldsymbol{\beta}_i\ \forall\ i$
2. $\boldsymbol{W} \mid \boldsymbol{b}, \boldsymbol{\beta}_i\ \forall\ i$
3. $\boldsymbol{\beta}_i\ \forall\ i \mid \boldsymbol{b},\boldsymbol{W}$

In order to take draws, one must know the posteriors for the mean and variance of a normal distribution, a multivariate normal distribution and an inverted Wishart distribution. Now, I will explain how to obtain draws from the three conditional distributions.

4.3.2 Taking draws from a multivariate normal distribution

The posterior distribution for a multivariate normal distribution with an unknown mean and known variance has to be determined. First, a one-dimensional distribution is made, then the results are generalized to a multivariate distribution.

Consider the random variable β, which is normal distributed with unknown mean b and known variance σ^2. In total, there are I realizations, labeled $\beta_i, i = 1, \dots, I$. The sample mean is derived by taking the average of all realizations: $\bar{\beta} = \frac{1}{I}\sum_i \beta_i$. Suppose, the market researcher's prior beliefs on the unknown mean are represented by a normal distribution with mean b_0 and variance s_0; $b \sim N(b_0, s_0)$. The posterior on b is then also normal distributed; $b \sim N(b_1, s_1)$, where

$$b_1 = \left(\frac{1}{s_0} b_0 + \frac{I}{\sigma} \bar{\beta}\right) s_1 \tag{4.13}$$

and

$$s_1 = \left(\frac{1}{s_0} + \frac{I}{\sigma}\right)^{-1}. \tag{4.14}$$

The mean of the posterior is the weighted average of the sample mean and the prior mean. If the number of realizations rises, the weight on the sample mean rises and the prior mean becomes less important. When the prior is nearly flat, the market researcher considers that all possible values of the parameters are equally likely. Then, we have a so-called "diffuse prior". When the variance of the prior, s_0, rises, the normal prior spreads out and becomes flatter. If $s_0 \rightarrow \infty$, then $\frac{1}{s_0} \rightarrow 0$ in equations (4.13) and (4.14) and the posterior distribution approaches $N(\bar{\beta}, \frac{\sigma}{I})$.

For the multivariate normal distribution, consider the K – dimensional random vector $\boldsymbol{\beta}$, which is multivariate normal distributed with unknown mean vector $\boldsymbol{b}$ and known covariance matrix $\boldsymbol{W}$. The market researcher observed a sample $\boldsymbol{\beta}_i, i = 1, \dots, I$ with sample mean $\bar{\boldsymbol{\beta}}$. The prior on $\boldsymbol{b}$ is multivariate normal: $\boldsymbol{b} \sim N(\boldsymbol{b}_0, \boldsymbol{W}_0)$. The posterior distribution is also multivariate normal:

$$N((\boldsymbol{W}_0^{-1} + I\boldsymbol{W}^{-1})^{-1}(\boldsymbol{W}_0\boldsymbol{b}_0 + I\boldsymbol{W}^{-1}\bar{\boldsymbol{\beta}}), (\boldsymbol{W}_0^{-1} + I\boldsymbol{W}^{-1})^{-1}).$$

In case there is a diffuse prior, the posterior approaches $N(\bar{\boldsymbol{\beta}}, \frac{1}{I}\boldsymbol{W})$.

If draws needed to be taken from this posterior distribution, the following should be done:

1. Choose $\boldsymbol{C}$ to be the Choleski factor of $\frac{1}{I}\boldsymbol{W}$.

2. Draw T I.I.D. standard normal deviates, $\zeta_t,\ t = 1, \dots, T$, and stack them in the vector $\boldsymbol{\zeta} = [\zeta_1, \dots, \zeta_T]$'.

3. Calculate $\widetilde{\boldsymbol{b}} = \overline{\boldsymbol{\beta}} + \boldsymbol{C\zeta}$.

The resulting vector $\widetilde{\boldsymbol{b}}$ is a draw from $N(\overline{\boldsymbol{\beta}}, \frac{1}{I}\boldsymbol{W})$.

4.3.3 Taking draws from an inverted Wishart distribution

The posterior distribution for an inverted Wishart distribution also has to be determined. First, there is a one-dimensional inverted Gamma distribution, then the results are generalized to an inverted Wishart distribution afterwards.

Consider a sample of I realizations, labeled $\beta_i, i = 1, \dots, I$. The sample variance equals $\bar{s} = \frac{1}{I-1}\sum_i(\beta_i - b)^2$. Suppose, the market researcher's prior beliefs on the unknown variance are represented by an inverted gamma distribution with shape parameter v_0 and scale parameter s_0; $\sigma \sim IG(v_0, s_0)$. Then, the posterior on σ is also inverted gamma distributed: $\sigma \sim IG(v_0, s_1)$, where

$$v_1 = v_0 + I \tag{4.15}$$

and

$$s_1 = \frac{v_0 s_0 + I\bar{s}}{v_0 + I}. \tag{4.16}$$

If the value of the shape parameter decreases, the prior becomes more diffuse. If $v_0 \to 1$, it is customary to set the scale parameter $s_0 = 1$. In case of this diffuse prior, the posterior distribution becomes $\sigma \sim IG\left(1 + I, \frac{1+I\bar{s}}{1+I}\right)$.

Consider the K – dimensional random vector $\boldsymbol{\beta}$, which is multivariate normal distributed with known mean vector $\boldsymbol{b}$ and unknown covariance matrix $\boldsymbol{W}$. The market researcher observed a sample $\boldsymbol{\beta}_i, i = 1, \dots, I$ with sample variance $\overline{\boldsymbol{S}} = \frac{1}{I}\sum_i(\boldsymbol{\beta}_i - \boldsymbol{b})(\boldsymbol{\beta}_i - \boldsymbol{b})'$. The prior on $\boldsymbol{W}$ is inverted Wishart with shape parameter v_0 and scale matrix $\boldsymbol{S}_0$; $\boldsymbol{W} \sim IW(v_0, \boldsymbol{S}_0)$. Then, the posterior distribution is also inverted Wishart: $\boldsymbol{W} \sim IW(v_1, \boldsymbol{S}_1)$, where v_1 is the same as in equation (4.15) and the formula for the scale matrix $\boldsymbol{S}_1$ is as follows:

$$\boldsymbol{S}_1 = \frac{v_0 \boldsymbol{S}_0 + I\bar{\boldsymbol{S}}}{v_0 + I}. \tag{4.17}$$

Similar to the one-dimensional case, the prior becomes more diffuse if the value of the shape parameter decreases. In order for the prior to integrate to one and have means, v_0 must exceed K. If the scale matrix $\boldsymbol{S}_1 = \boldsymbol{E}$, where $\boldsymbol{E}$ is the K – dimensional identity matrix, then the posterior distribution becomes $\boldsymbol{W} \sim IW\left(K + I, \frac{K\boldsymbol{E} + I\bar{\boldsymbol{S}}}{K+I}\right)$.

If the researcher wants to take draws from this posterior distribution, he should do the following:

1. Take v_1 draws of K – dimensional vectors, whose elements are independent standard normal deviates. These draws are labeled $\boldsymbol{\zeta}_t$, $t = 1, \ldots, v_1$.

2. Choose $\boldsymbol{D}$ to be the Choleski factor of the inverse of $\boldsymbol{S}_1$; $\boldsymbol{D}\boldsymbol{D}' = {\boldsymbol{S}_1}^{-1}$.

3. Calculate the variance of draws $\boldsymbol{Z}$ from a distribution with variance $\boldsymbol{D}\boldsymbol{D}' = {\boldsymbol{S}_1}^{-1}$,

$$\boldsymbol{Z} = \frac{1}{v_1} \sum_t (\boldsymbol{D}\boldsymbol{\zeta}_t)(\boldsymbol{D}\boldsymbol{\zeta}_t)'. \tag{4.18}$$

4. Take the inverse of $\boldsymbol{Z}$.

The matrix $\tilde{\boldsymbol{S}} = \boldsymbol{Z}^{-1}$ is a draw from $IW(v_1, \boldsymbol{S}_1)$.

4.3.4 Taking draws with the Metropolis – Hastings algorithm

The posterior for each respondent's $\boldsymbol{\beta}_i$, conditional on the choices they made and the population mean and variance of $\boldsymbol{\beta}_i$, is as follows:

$$\pi(\boldsymbol{\beta}_i | \boldsymbol{b}, \boldsymbol{W}, \boldsymbol{Y}_i) \propto L(\boldsymbol{y}_i \mid \boldsymbol{\beta}_i)\phi(\boldsymbol{\beta}_i \mid \boldsymbol{b}, \boldsymbol{W}). \tag{4.19}$$

It is not easy to take draws from this posterior distribution. Therefore, the Metropolis – Hastings algorithm is used. The procedure to take draws is explained in seven consecutive steps:

1. Start with $\boldsymbol{\beta}_i^0$.

2. Draw K independent values from a standard normal distribution. The draws are placed into the vector $\boldsymbol{\omega}^1$.

3. A trial value for $\boldsymbol{\beta}_i^1$ is created: $\widetilde{\boldsymbol{\beta}_i^1} = \boldsymbol{\beta}_i^0 + \lambda \boldsymbol{G}\boldsymbol{\omega}^1$, where λ is a scalar specified by the market researcher and $\boldsymbol{G}$ is the Choleski factor of $\boldsymbol{W}$.

4. Draw a standard uniform variable μ^1.

5. The 'likelihood' that $\widetilde{\boldsymbol{\beta}_i^1}$ seems to be a more accurate estimate than $\boldsymbol{\beta}_i^0$ is based on the following ratio: $Q = \frac{L(\boldsymbol{y}_i|\widetilde{\boldsymbol{\beta}_i^1})\phi\left(\widetilde{\boldsymbol{\beta}_i^1}|\boldsymbol{b},\boldsymbol{W}\right)}{L(\boldsymbol{y}_i|\boldsymbol{\beta}_i^0)\phi\left(\boldsymbol{\beta}_i^0|\boldsymbol{b},\boldsymbol{W}\right)}$.

6. If $\mu^1 \leq Q$, then $\widetilde{\boldsymbol{\beta}_i^1}$ is accepted and $\boldsymbol{\beta}_i^1 = \widetilde{\boldsymbol{\beta}_i^1}$. Otherwise, $\widetilde{\boldsymbol{\beta}_i^1}$ is rejected and $\boldsymbol{\beta}_i^1 = \boldsymbol{\beta}_i^0$.

7. Repeat steps 1 – 6 many times. For large enough values of h, $\boldsymbol{\beta}_i^h$ is a draw from the posterior distribution in equation (4.19).

The speed of convergence depends on the scalar λ in the third step. This scalar determines the step size. Steps that are too large are not likely to be accepted. Smaller steps imply that the Metropolis – Hastings algorithm takes more iterations in order to reach convergence and embody more serial correlation in the draws after convergence. The optimal acceptance rate is around 0.44 in the one-dimensional case and it declines to 0.23 for high dimensions (Gelman *et al.*, 1994). The initial value of λ can be set by the market researcher to achieve an acceptance rate in the preferred neighborhood.

During the iterative process, the value can be adjusted. In each iteration, a trial $\boldsymbol{\beta}_i$ is accepted or rejected. If in an iteration the acceptance rate among the I observations is above a given predefined threshold value, then the value of λ must be raised. If the acceptance rate is below this threshold value, then the value of λ must be lowered. This ensures that the value of λ is close to the specified value during the iteration process.

As it is now clear how to obtain draws from the three conditional distributions, the procedures are combined into a MCMC procedure for the three sets of parameters. The estimation procedure can be written in a more concise form. One should start with any initial values $\boldsymbol{b}^0, \boldsymbol{W}^0$ and $\boldsymbol{\beta}_i^0 \ \forall\, i$. The h-th iteration of the MCMC procedure consists of the following three steps:

1. Draw $\boldsymbol{b}^h$ from $N(\overline{\boldsymbol{\beta}}^{h-1}, \frac{1}{I}\boldsymbol{W}^{h-1})$, where $\overline{\boldsymbol{\beta}}^{h-1}$ is the mean of the $\overline{\boldsymbol{\beta}}_i^{h-1}$'s.

2. Draw $\boldsymbol{W}^h$ from $IW\left(K + I, \frac{K\boldsymbol{E}+IS^{h-1}}{K+I}\right)$, where $S^{h-1} = \frac{\boldsymbol{1}}{\boldsymbol{I}}\sum_i\left(\boldsymbol{\beta}_i^{h-1} - \boldsymbol{b}^h\right)\left(\boldsymbol{\beta}_i^{h-1} - \boldsymbol{b}^h\right)'$.

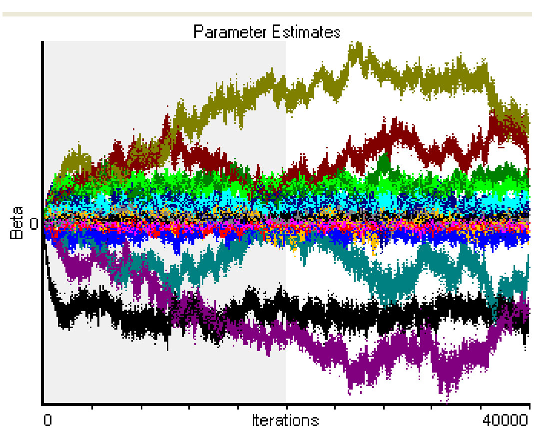
3. For each i, draw $\boldsymbol{\beta}_i^h$ using one iteration of the Metropolis – Hastings algorithm, starting from $\boldsymbol{\beta}_i^{h-1}$ and using the normal density function $\phi(\boldsymbol{\beta}_i | \, \boldsymbol{b}^h, \boldsymbol{W}^h)$.

These three steps are repeated many times until convergence is reached The resulting values converge to draws from the joint posterior distribution of $\boldsymbol{b}, \boldsymbol{W}$ and $\boldsymbol{\beta}_i \; \forall \, i$. Then, the mean and standard deviation of the draws can be calculated, which leads to the estimates and standard errors of the parameters. However, it first has to be verified whether convergence has been achieved.

4.3.5 Checking for convergence

Convergence is not always easy to determine (Kass *et al.*, 1998). One of the options a market researcher can choose is to start the MCMC procedure from many different points and to test if the posterior mean is the same when calculated from each sequence generated from the different starting points (Gelman and Rubin, 1992). A market researcher should first specify a number of iterations. The total number of iterations is split in two sets. These sets do not necessarily have the same number of iterations, although this is common in practice. The first set consists of iterations prior to assuming convergence, also called ‘burn in’. These iterations are used to observe if there is trending behavior. Convergence has been achieved when the draws do not deviate much from the posterior anymore. The second set of iterations consists of subsequent draws used to obtain estimates and standard errors of the parameters.

To illustrate this, two figures are presented below that provide an impression of how a market researcher can observe whether convergence has been achieved or not. In figure 4.2a, convergence has not been achieved, because after ‘burn in’ (grey region; until 20.000 iterations) at least four sequences do not move towards the mass of the posterior. The most obvious solution is to use more iterations before assuming convergence. In figure 4.2b, the pattern after ‘burn in’ is similar to its pattern before. So, convergence has been reached.



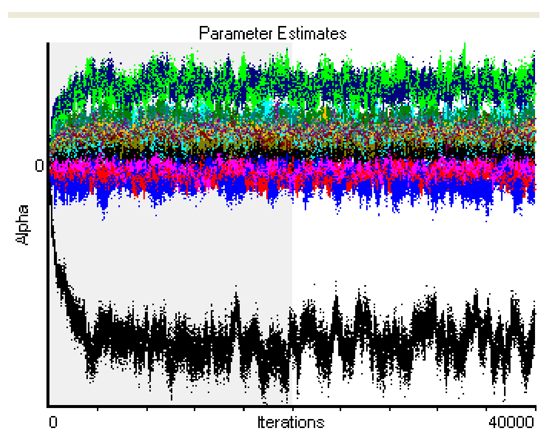


Figure 4.2a-b: No convergence reached versus convergence reached

The successive draws are not independent. Even after convergence, each iteration is built on the previous one. Moreover, for the Metropolis – Hastings algorithm, in roughly $(1 - \lambda) * 100\%$ of all iterations, the value for the next iteration remains the same. The dependence among successive draws leads to a decrease in precision for the means and standard errors. Thinning can compensate for the dependence by reducing the amount of correlation among the draws. A thinning factor of τ means that results are only retained for each τ-th iteration and others are discarded. For example, if the total number of draws that should be saved is 1,000 and $\tau = 10$, the market researcher should choose 10,000 iterations after burn-in. Then, the number of iterations for burn-in is often 10,000 as well.

There are certain statistics indicating 'goodness of fit' that are useful in assessing convergence. The measures used here are derived from the likelihood of the data. As this is a very small number, the log likelihood of the data should be used.

The first measure is the 'percent certainty'. This indicates how much better the solution is compared to chance (Hauser and Glen, 1979). Percent certainty equals the final log likelihood minus the log likelihood of a chance model, divided by the negative of the log likelihood of a chance model. The value lies in the interval [0, 1], where 0 means that the model fits the data at only the chance level and 1 means there is a perfect fit. For example, a value of 0.70 means that the log likelihood is 70% of the way between the expected value by chance and the value for the perfect fit.

The second measure is the 'root likelihood'(RLH). For the RLH, the measure of goodness of fit is obtained in a similar way. The RLH for each respondent is calculated as follows: the posterior means of a respondent's part-worths are used in the mixed logit model to estimate the probability of each of the choices made in all the choice tasks by the respondent, and the geometric mean of these probabilities must be calculated (Wonder *et al.*, 2008). The model RLH is calculated as the arithmetic average of all the respondent RLH values. For example, if there are three alternatives in each choice task and there is no information about part-worths, the typical prediction that can be made about the probability is that all the choices that are predicted correctly is 1/3. Thus, the expected model RLH is 1/3 as well. The actual value for the RLH also lies in the interval [0, 1], but the value is interpreted in another way as for the percent certainty. If the actual value for the RLH is 0,70, this means the fit is 0,70/(1/3) = 2,1 times better than the chance level.

As iterations progress, both percent certainty and RLH tend to increase first and at a certain time level off. Thereafter, they oscillate randomly around their final values. This can be seen as a proof for convergence. However, convergence cannot be identified until long after it has occurred. Therefore, planning a large number of iterations and check for stability is the key. As both measures are similar, the choice between them depends on the market researcher's preference.

4.4 The effect of using constraints

Choice-based conjoint studies often include certain product attributes for which almost every respondent is expected to prefer a certain level to another. However, the estimated individual part worths do not always follow this expected order. This can lead to several problems, because part worths with the wrong slopes and coefficients with the wrong signs are causing the results to make less or even no sense at all. The solution for this problem is to use constraints on the order of the part-worths within attributes and on signs of linear coefficients.

The constraints are the same for each respondent. They should only be used for attributes with an unambiguous a-priori preference order (e.g. volume, quality). For example, using the constraints on the order of the part-worths for the attribute 'price' should be done with caution. A lower price is not always preferred (e.g. sometimes respondents prefer a higher price, because they think this means a higher products quality). Constraints can be very valuable for a market researcher who is interested in the prediction of individual choices.

Constraints are expected to reduce variance and increase bias (Wittink, 2000). Hit rates are sensitive to both variance and bias. By trading a large amount of variance for a small amount of bias, the hit rates are likely to be improved. For aggregated measures, the variance is relatively small, because the random error is likely to be averaged out. Therefore, aggregate share predictions are affected most by the bias. So, if the constraints are going to be implemented, there will be more realistic results at the cost of a decrease in accuracy.

5. Empirical application on credit cards – Data

Standard choice-based conjoint studies are used to describe preferences for a product consisting of different attributes. The effect of visualization is usually ignored. However, brands use different communication styles to show the characteristics of the product. This makes it harder for consumers to directly compare offers and make the best decision. This chapter describes the data that is used to quantify how the visual attribute representations affect consumers' preferences in the market of credit cards.

Section 5.1 motivates why the study is executed in the field of credit cards. Also, a detailed description of the product attributes is presented. Section 5.2 specifies which visualizations are applied to the product attributes for each of the three visualization conditions. In section 5.3, the methodology for this study is presented. A specific study structure is used to obtain the respondent's data. Not every respondent gets to answer the same questionnaire. Every respondent is only exposed to a standard CBC exercise and one CBC exercise where one of the three visualization conditions is used. The study structure is described in detail in section 5.4. In section 5.5, several software programs are compared to show which one is the best to use for this type of study.

5.1 The important attributes for the choice of a credit card

The reasons why the study concerns credit cards, is because credit cards are important products (especially in the United States of America, where almost every adult has at least one credit card), everybody knows what the products are about, credit cards are offered by many competitors using various communication styles, and because credit cards are popular.

The choice for a certain credit card has an impact on the buyer's financials for a long time. Therefore, the buyer should look for the best available offer. That is not an easy exercise, because a credit card offer is described by various attributes (e.g. annual percentage rate, bonuses, transfer fees) and not every consumer is completely aware of the consequences of a particular attribute level.

Credit cards are offered by many banks. Each bank uses different communication styles and this makes it difficult to directly compare the offers. For example, see the three different credit card offers of JP Morgan Chase, Capital One and Bank of America in figure 5.1.

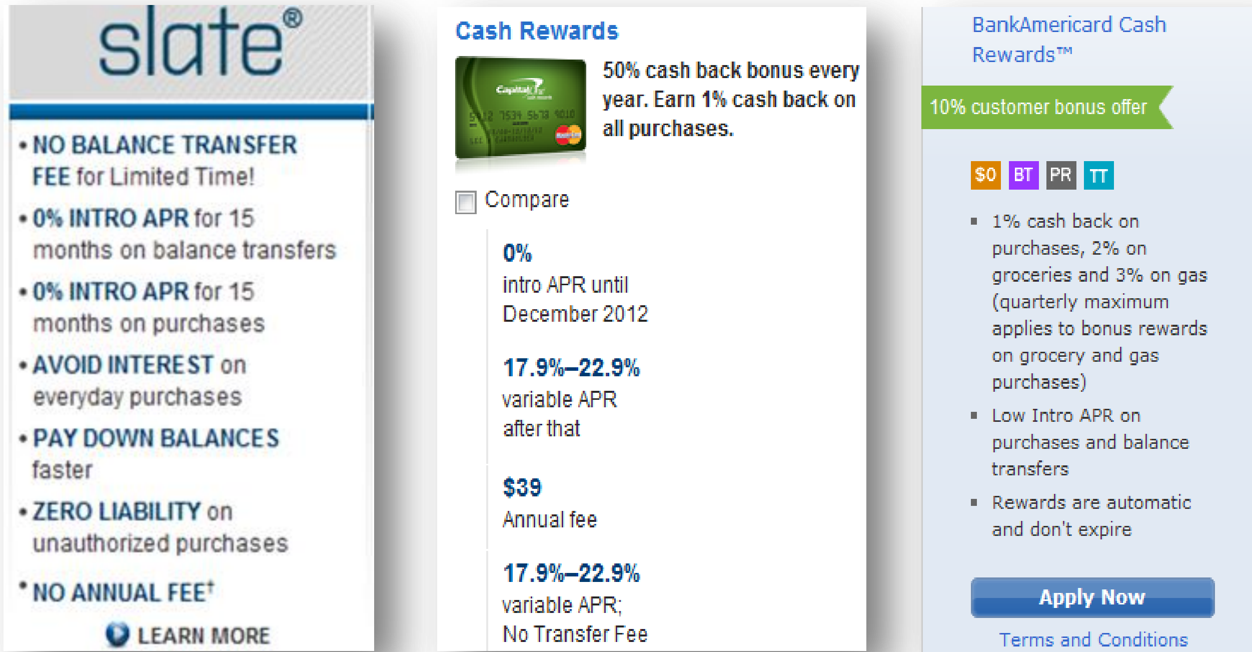
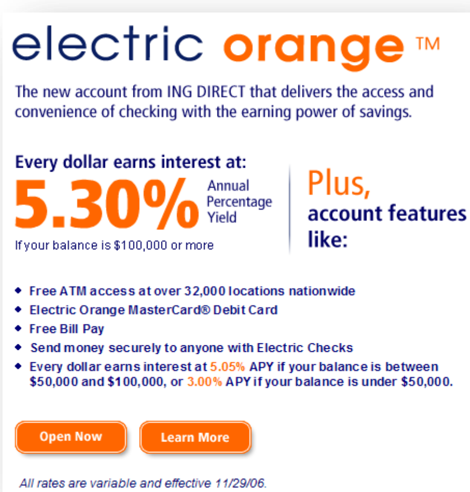


Figure 5.1a-c: Credit card offers of JP Morgan Chase, Capital One and Bank of America

In theory, one possibility to overcome the different communication styles is by not paying attention to it at all and only focus on the text in the offers. In figure 5.2, two offers are displayed: one with visualization techniques applied to it and one without any visualization techniques. Although both concepts describe the same product, consumers are not likely to react in the same way to these offers. So, the validity of ignoring the visualization can be low. Being good at modeling the likelihood of choice of the second offer could be irrelevant and potentially misleading, because it could undermine the impact of visualization on choice behavior.



≠

Credit card: Electric Orange by ING Direct

- Annual interest yield:
  - 3.00% if account balance is up to $50,000
  - 5.05% if account balance is between $50,000 and $100,000
  - 5.30% if account balance is over $100,000
- ATM access: free at 32,000 locations
- Bill pay: free of charge
- Electric checks: included
- Debit card: included

Figure 5.2a-b: Credit card offers with and without visualization techniques

A simplification technique that can be applied is to ask the respondents to choose between product concepts that show product information in a consistent and uniform way, keeping attributes in the same position and using similar, neutral phrasing of the attribute levels. If visualization effects are included in the research designs, more realistic choice tasks are created, potentially driving higher external validity. Before turning to the visualization effects for the product concepts, I describe the important attributes that are included in the product concepts.

In total, six attributes were chosen. The choice for the six attributes was based on the relevance for the consumer and on the frequency with which the attributes occurred in the various offers from the banks. SKIM started with a broad selection of attribute levels. The first four attributes are considered to be the most important attributes, and so these are called ‘primary attributes’ (see table 5.1). The remaining two attributes are called ‘secondary attributes’ and are displayed in table 5.2. The constraints that are used for the preference order of the different attribute levels are presented in the tables as well.

Table 5.1: Primary product attributes

Attribute	Level
Annual membership fee	• $0.00 • $40.00 • $67.50 • $95.00 *Constraint: $0.00 > $40.00 > $67.50 > $95.00*
Waived fee	• No annual membership fee for six months, then $x • $x annual membership fee
Special benefit	• No special benefit • Special benefit: Unlimited 1% Cash Back • Special benefit: $200 Cash Back after you spend $500 in 3 months • Special benefit: Earn 2 Air Miles for each $1 spent *Constraint: Special benefit > No special benefit ∃ Special benefit*
Annual percentage rate (APR)	• 0% Intro APR for 6 months on purchases; after that a variable APR of 10.99% - 17.99% depending on your credit worth (“level 1”) • 0% Intro APR for 12 months on purchases; after that a variable APR of 10.99% - 17.99% depending on your credit worth (“level 2”) • 0% Intro APR for 12 months on purchases; after that a variable APR of 12.99% - 22.99% depending on your credit worth (“level 3”) • 0% APR • 15% APR • 20% APR *Constraints: “level 2” > “level 1” & “level 2” > “level 3”* *Constraint: 0% APR > 15% APR > 20% APR*

Table 5.2: Secondary product attributes

Attribute	Level
Balance transfers/cash advance fees	• Balance Transfers: Either \$5 or 3% of the amount of each transfer whichever is greater. Cash Advances: Either \$10 or 3% of the amount of each transaction whichever is greater. ("level 1") • Balance Transfers: Either \$10 or 5% of the amount of each transfer whichever is greater. Cash Advances: Either \$10 or 4% of the amount of each transaction whichever is greater. ("level 2") *Constraint: "level 1" > "level 2"*
Late payment fees	• Late Payment: Up to \$15 if the balance is less than \$150; Up to \$45 otherwise. Return Payment: Up to \$35. • Late Payment: Up to \$35. Return Payment: Up to \$25. • Late Payment: Up to \$29 if the balance is less than \$1000; Up to \$39 otherwise. Return Payment: Up to \$23.

The annual membership fee is one of the most important attributes, because it shows the direct price consumers have to pay for the product on a yearly basis. Some banks offer credit cards that have no annual membership fee and others require fees of almost \$95 each year. The average yearly membership fee is in the neighborhood of \$40 and this attribute level is included as well. The fourth attribute level is chosen to be exactly in between the average and the highest fee. It is possible that the membership fee only has to be paid when you are a member for at least six months. This is represented by the second attribute.

The market for credit cards faces strong competition. There are so many offers and banks hope to stand out and attract potential consumers by offering special benefits. As special benefits are used to attract and delight consumers, they are often displayed more prominently. Typical examples are unlimited x% cash back or saving points in exchange for gifts (e.g. Air Miles). The last primary attribute is the annual percentage rate (APR). This can be a fixed number or it can depend on the user's credit worth. Similar to the membership fees, some credit card users do not have to pay APR in the first six months. However, the APR can be relatively high afterwards.

The balance transfers/cash advance fees and the late payment fees are variable penalty fees the user can be charged for when using a credit card. A credit card balance transfer is the transfer of the money owed in a credit card account to an account held at another credit card company. The user pays a fixed amount or a fixed percentage of the balance transfer, whichever is greater. Late payment fees are fines the user should pay if he has not paid on time.

5.2 The three visualization techniques applied to credit card offers

The effect of attribute level visualization on consumer choice behavior is explored for three different conditions: pop-ups and footnotes, visible or hidden attributes and font size variations. In the subsections 5.2.1 – 5.2.3, a description of each of these visualization techniques is presented. Also, examples of choice tasks are provided in order to make clear what the visualization techniques actually look like for the product concepts shown to the respondent.

5.2.1 Pop-ups and footnotes

Instead of presenting attribute levels using plain and descriptive text, concise and captivating text is used to cover the most important information. The remaining information is given in a pop-up or footnote. For the pop-up, the user has to click on a specific button to view the full explanation. Footnotes are presented at the bottom, where the full details can be read. In figure 5.3, a choice task is shown where pop-ups and footnotes are presented within the three product concepts.

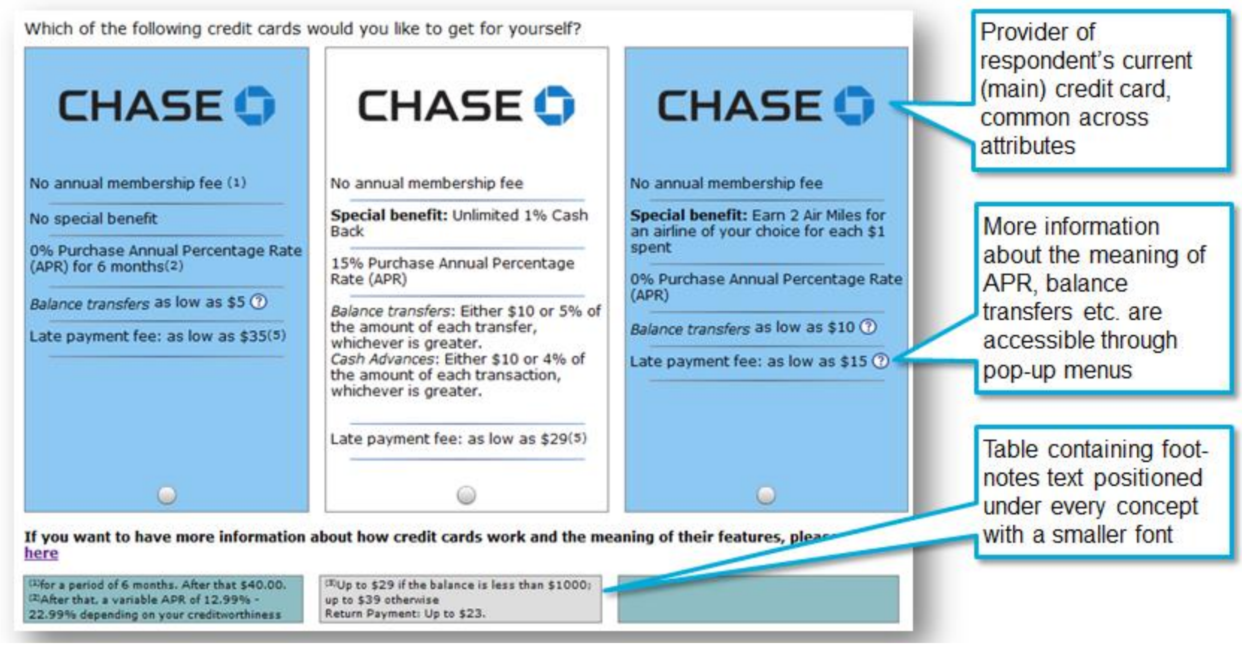


Figure 5.3a-c: Pop-ups and footnotes

Several models are used to test whether this visualization technique affects consumer choices. In the null/base model, the product attributes are included without using any visualization technique. For the full model, the main effects for four attributes (annual membership fee, APR, balance transfers and late payment fees) are included. The single effect model only includes the main effect for the most important attribute (annual membership fee). Finally, one can test whether including a first-order interaction for the first two primary attributes within the single effect model leads to higher validity. To test which model has the highest validity, the hit rates on hold-out tasks are compared. A higher hit rate implies a higher validity.

5.2.2 Visible or hidden attributes

Instead of always making attribute levels fully visible, some were hidden, making them only accessible through a “General terms and conditions” window. For hidden attributes, the attribute level description is not present in the main concept. It is available only in a pop-up containing the full product conditions. Figure 5.4 shows how the hidden, but accessible, attribute levels are presented within the product concepts of a choice task.

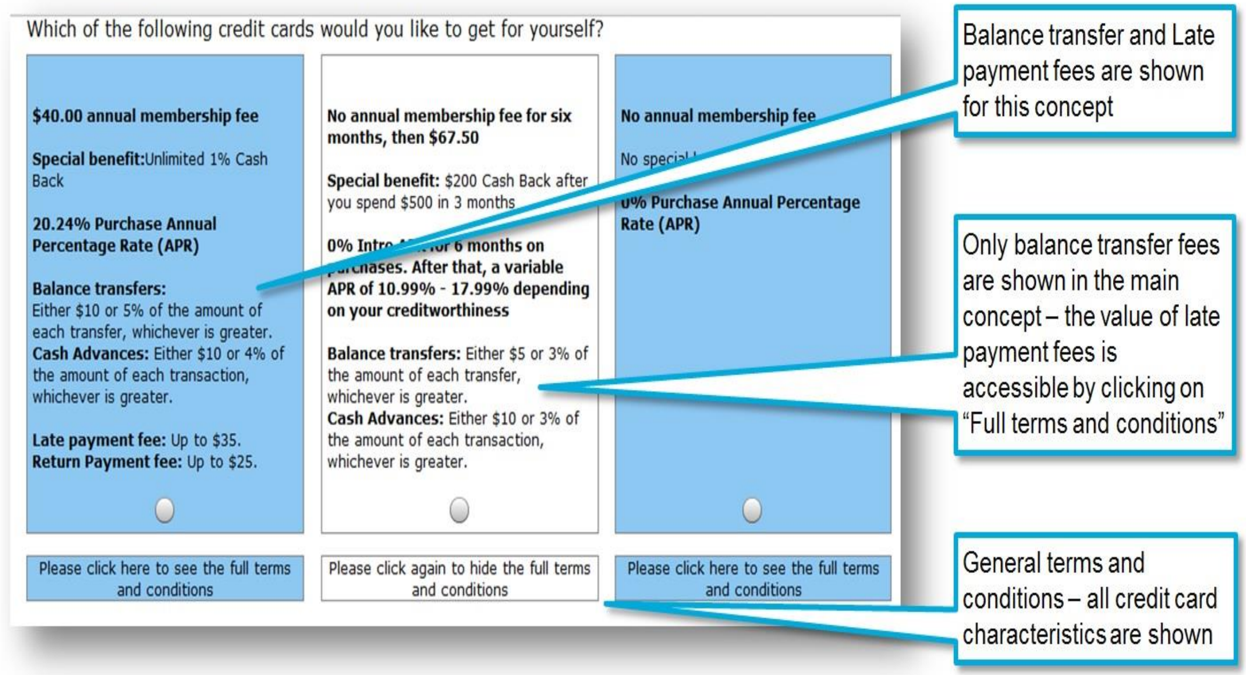


Figure 5.4a-b: Visible or hidden attributes

Most of the consumers pay close attention to at least one of the primary attributes when they select a credit card. Therefore, banks rarely hide the primary attributes. For this study, only the secondary attributes can be hidden and made visible through a link. Hiding secondary attributes is not necessary to improve validity and to have an impact on choice behavior, just because the attributes are already considered to be less important. However, this does not always hold. A consumer is less likely to trust a credit card offer when there is an absolute lack of information about certain characteristics. Another possibility is that the offer is more likely to be chosen, because it is less cluttered.

Two different models are used. In the null/base model, the product attributes are included without using any visualization techniques. Therefore, this model functions as a control model. The second model is the main effects model, which includes the visualization techniques for the secondary product attributes. As the visualization technique is only applied to the two less important attributes, models where the visualization technique is applied to only one of these attributes are not considered.

5.2.3 Font size variations

Instead of presenting all attribute levels using the same font size, the font sizes were varied to emphasize or play down levels or aspects of levels. Font sizes highlight or downplay attributes. The most important attributes for the consumers searching for a credit card can be highlighted using a larger font. Attributes that are less important can be displayed less prominently using a smaller font. The annual membership fee, special benefit and APR are shown with a large or normal font size (independently from the content). The balance transfers and late payment fees are shown with a normal or small font size (independently from the content). Figure 5.5 shows how the different font sizes are presented within three product concepts of a choice task.

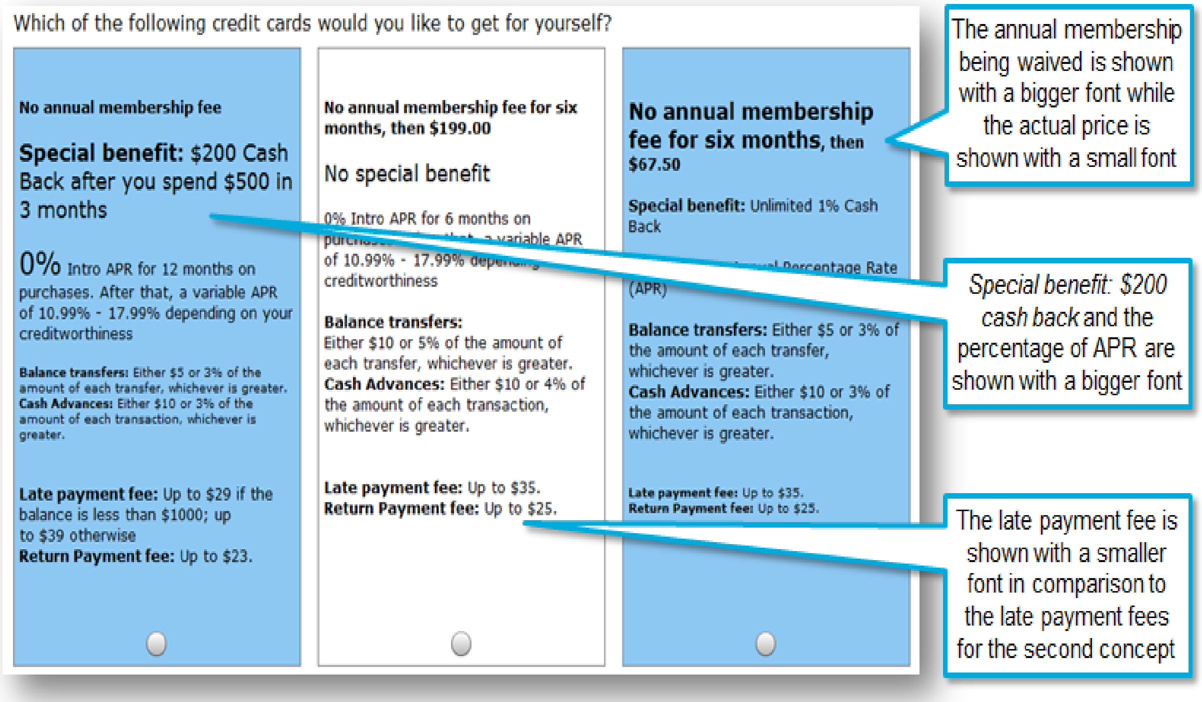


Figure 5.5a-b: Font size variations

Font size variations give the consumer an easy guidance through all the features of the bank's offer. It is expected that consumers focus more on the attribute levels that are highlighted, because a large font means more visibility. However, the impact of the visualization can depend on the level content. One way to test this is to measure how often the bigger font size is preferred over the smaller font size for each product attribute.

Three different models are used. In the null model, the product attributes are included without using any visualization technique. The second model is the main effects model, which includes the visualization techniques for all product attributes. The third model tests whether including a first-order interaction for the first two primary attributes leads to higher validity. Again, a higher hit rate implies a higher validity.

5.3 The applied methods to test the impact of visualization techniques

The main goal of this thesis is to quantify the effects of visualization on the actual choices of the respondents. The mixed logit model, which is explained in the previous chapter, is used for this study. The parameter vector includes parameters that can be used to study the effects of visualization. The vector $\boldsymbol{\beta}_i$ has all the utilities for respondent i, which are the values describing the attractiveness of the levels of the product attributes, but also the values describing the attractiveness of a particular visualization effect. For the sake of simplicity, only one type of visualization technique is used at a time. The visualization effects can be retrieved by distinguishing between the effects of the levels of the product attributes and the visualization effects. The parameter vector $\boldsymbol{\beta}_i$ can be seen as the union of two independent subsets of parameters: $\boldsymbol{\beta}_i = (\boldsymbol{\gamma}_i;\ \boldsymbol{\delta}_i)$, where $\boldsymbol{\gamma}_i$ contains the values describing the attractiveness of the levels of the product attributes and $\boldsymbol{\delta}_i$ contains the values describing the effects of using visualization techniques.

The parameters for both $\boldsymbol{\gamma}_i$ and $\boldsymbol{\delta}_i$ can only be obtained if the vector of independent variables is recoded correctly. In section 3.1, a procedure is described that is used to obtain a dummy vector, where a '1' corresponds to a level of a product attribute being present and '0' otherwise. This is needed for the estimation of the vector $\boldsymbol{\gamma}_i$. The parameters belonging to $\boldsymbol{\delta}_i$ can be estimated only if the vector of explanatory variables (recoded as a vector with 0/1 values) is extended with entries that indicate whether a specific visualization effect for the level of an attribute is present or not.

For example, let us consider a credit card offer where pop-ups and footnotes are used to captivate the text and cover the most important information for the attributes 'Special benefit' and 'late payment fees'. The attribute levels are:

- $40.00
- $x annual membership fee
- Special benefit: Unlimited 1% Cash Back
- 15% APR
- Either $5 or 3% of the amount of each transfer whichever is greater…
- Late Payment: Up to $35. Return Payment: Up to $25

The corresponding dummy vector is $\boldsymbol{\gamma}_i = [0,1,0,0,0,1,0,0,0,0,0,1,1,0,1]'$. The usage of pop-ups and footnotes must be incorporated in the dummy vector $\boldsymbol{\delta}_i$. The vector $\boldsymbol{\delta}_i$ contains the same number of entries as attributes; in this case six elements. Only the third and sixth additional entries get the value '1', because a visualization technique is used for these attribute only. The new dummy vector $\boldsymbol{\beta}_i = (\boldsymbol{\gamma}_i;\ \boldsymbol{\delta}_i)$ becomes: [0,1,0,0,0,1,0,0,0,0,0,1,1,0,1,0,0,1,0,0,1]'.

Visualization techniques are often applied to a limited number of attributes. This makes sense, because using the same visualization technique for every product attribute does not lead to a specific attribute standing out anymore.

5.4 The structure of the study

The CBC study conducted at SKIM resulted in available information for 1.665 respondents from the United States of America. The questionnaire that was used to gather the data is presented in appendix A1. This questionnaire was built by SKIM. Each respondent, after being screened as qualified for the study, was exposed to only one of the three visualization conditions. Otherwise, the effort required from the respondents to participate would have become too big. An almost equal distribution of respondents over the three visualization conditions was obtained. The overall study structure is presented in figure 5.6.

The CBC exercise with a particular visualization condition contained eight choice tasks and each choice task contained three product concepts. The last choice task was used as a hold-out task. The product concepts were created according to a balanced incomplete block design (Bose, 1942). This is abbreviated to (n, b, r, p, λ)-BIBD, where n is the number of product concepts, b the number of choice tasks, r the number of repetitions of a product concept over all blocks, p the number of product concepts in each choice task and λ the number of times a pair of product concepts is included. The parameters of the BIBD are related. In order to obtain a BIBD, the following equations have to hold:

$$n * r = b * p, \tag{5.1}$$

$$r * (p - 1) = \lambda(n - 1). \tag{5.2}$$

The BIBD allows the market researcher to test the preference for the different product attributes in an efficient way, because each interesting pair of attribute levels, but also the attribute levels themselves, are only used a limited but sufficient number of times. In order to have a reference measurement of the respondents' preferences, each respondent answered eight similar choice tasks for a plain CBC as well. In the plain CBC, the product concepts in each choice task were fully described without any visualization effects.

Differences in the results can occur for respondents who saw visualization effects before or after having seen the plain choice tasks. Showing the control exercise first can cause a learning effect. For example, if respondents have seen the plain tasks first they may give annual membership fees more or less attention when they are presented with a larger font size in comparison to the case in which the respondents have seen no plain task before. The most realistic situation according to me is first seeing the exercise with visualization effects.

Also, the respondents were asked to answer demographic and attitude questions. The answers from these questions can be used to check to which group of consumers the respondent belongs to. However, including these results is beyond the scope of this thesis.

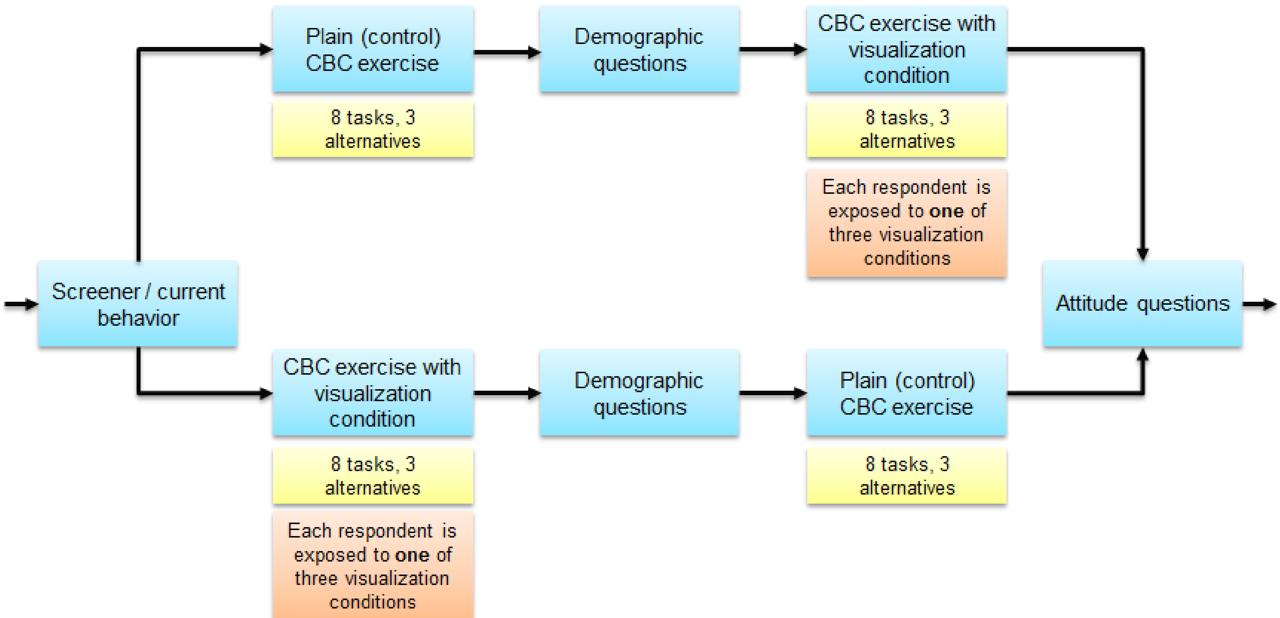


Figure 5.6: Structure of the study

5.5 Software

There are different software programs that can be used for conjoint studies. The software that was used for this study is Sawtooth Software's CBC/HB and R. Another software program suitable for conjoint studies is Winbugs.

My motivation to work with CBC/HB is because it is a user-friendly program that runs significantly faster than the other software programs. The disadvantage of working with CBC/HB is that you need an expensive license to have all the benefits of using this program optimally. To overcome paying for a good program for conjoint studies, R can be considered. R is an open source language program that needs free of charge packages to run specific analyses. There are already some valuable packages available for Bayesian analysis of conjoint data, such as "bayesm", that can be modified to make it work for a particular CBC study. The script code in R used for this study is described in appendix A2. WinBugs is a free software program as well, but it is less user-friendly and takes more time to run simulations. The results for CBC/HB and R are very similar and are described in chapter 6.

6. Empirical application on credit cards – Results

In the credit card market, different communication styles are used to show the characteristics of the products. This makes it harder for consumers to directly compare the offers and to make the best decision. In the previous chapter, the most important attributes for credit cards are described and the three different visualization techniques are explained in detail. This chapter quantifies how the visualization of choice options impact the actual choice for the empirical application on credit cards.

Section 6.1 describes if using concise and captivating descriptions by means of pop-ups and footnotes affect consumer choices. In section 6.2, the results for the second visualization technique are described. The impact on the actual choice of hiding secondary attributes from the concept are quantified by placing these in a 'General terms and conditions' window. The effect of using variable font sizes to highlight or downplay aspects on the actual choice is described in section 6.3. Only data from respondents who completed the survey is used.

6.1 Results for pop ups and footnotes

For the plain CBC exercise, all information about the product concepts is readily available on screen. For product concepts consisting of several product attributes, this can lead to too much information on screen. Credit card companies often want to focus more on the important aspects of the offer. One straightforward solution is to use pop-ups and footnotes to show concise and captivating descriptions of the attribute levels. More information about the attribute levels is available in the pop-ups and footnotes.

In subsection 5.2.1, four different models are described:

1. Full model: main effect for four product attributes (annual membership fee, APR, balance transfers and late payment fees).

2. Single effect model: main effect for the most important attribute (annual membership fee).

3. Single effect model with interaction: first-order interaction for the most important attribute (waived annual membership fee).

4. Null model: main effects for visualization technique not included.

The average part-worth utilities for all the models are described in appendix 3. To test validity, the hit rates on hold-out tasks are compared. The hit rates for these models are presented in table 6.1. The hit rates are based on input from 557 respondents.

Table 6.1: Hit rates for models with pup-ups and footnotes

Model	Model description	Hit rate
Full model	Main effect for four attributes	59%
Single effect model	Main effect for most important attribute	62%
Single effect model with interaction	First-order interaction for most important attribute	65%
Null model	No visualization technique applied	38%

The hit rates show that without any visualization technique, the hit rate is nearly 1/3, which equals the probability that one out of three concepts is chosen at random. If pop-ups and footnotes are applied to four attributes, the hit rate becomes 59%. This means that including pop-ups and footnotes affect consumer choices. The hit rate is 3% higher if the specific visualization technique is only applied to the most import attribute: annual membership fee. So, actual consumer choices are driven more by applying the visualization technique only for the most important attribute. Finally, by allowing a first-order interaction for the yearly membership fee to be waived or not, the hit rate is 65%.

The task order (plain CBC exercise or CBC exercise with pop-ups and footnotes first) had no significant effect on hold-out performance. This was detected by running two independent estimation processes.

Annual membership fee is considered to be the most important attribute and this is proven below. For each respondent, the importance of each product attribute is measured by the distance between the highest and lowest part-worth utilities belonging to its attribute levels. The attribute importance can be calculated by dividing this number, representing the distance, by the sum of all the distances for all product attributes. In table 6.2, it is illustrated how the importance of each attribute is calculated for the first respondent of the plain CBC exercise.

Table 6.2: Calculating attribute importance

Product attribute	Part-worth utilities						Max - min	Importance
Annual membership fee	7,81	-7,41	-2,53	2,13			15,22	73%
Waived fee	0,54	-0,54					1,07	5%
Special benefit	-1,28	0,80	0,45	0,03			2,08	10%
Annual percentage rate	-0,24	0,72	-0,11	-0,82	-0,09	0,54	1,53	7%
Balance transfers	0,06	-0,06					0,12	1%
Late payment fees	-0,44	0,48	-0,04				0,93	4%

The overall attribute importance is obtained by taking the average for the attribute importance over all respondents. The results are presented in table 6.3. Annual membership fee is the most important product attribute, with a 52% attribute importance.

Table 6.3: Overall attribute importance

Product attribute	Attribute importance
Annual membership fee	52%
Waived fee	4%
Special benefit	17%
Annual percentage rate	21%
Balance transfers	1%
Late payment fees	5%

The choice for the best model depends on two characteristics: the hit rate and the number of parameters. A higher hit rate means a better external validity. A model with a relatively low number of parameters to estimate is running faster. So, the preferred model has a high hit rate and a relatively low number of parameters. The full model contains more parameters than the single effect models and has a lower hit rate. The hit rate of the null model is low compared to the single effect models. Thus, the best choice is the single effect model with or without interaction. Including interaction in the single effect model increases the number of parameters with two and results in a 3% gain in hit rate. Therefore, the choice between them depends on the market researcher's preferences. In this thesis, it is chosen to continue with the single effect model without interaction.

Using pop-ups and footnotes affect consumer choices. It makes consumers trade to higher annual membership fees. This can be seen in tables 6.3 en 6.4, where the share of first choice is given for several product concepts. Only the annual membership fee is different; attribute levels not listed are the same across the credit card offers (ceteris paribus). The higher annual membership fees are preferred if a pop-up is used. The results are very interesting, because this presents an insight how easy it is to affect consumer choices and how big the impact can be from both a consumer and company perspective.

Table 6.4: Trade to higher annual fees (1)

Concept characteristics	Share
$40 annual membership fee (plain)	16%
$67 annual membership fee (pop-up)	76%
$67 annual membership fee (footnote)	8%

Table 6.5: Trade to higher annual fees (2)

Concept characteristics	Share
$40 annual membership fee (plain)	48%
$95 annual membership fee (pop-up)	50%
$95 annual membership fee (footnote)	2%

It would have been interesting to know if combining both pop-ups and footnotes for one product concept would lead to a bigger impact on actual choice, compared to only using the same amount of pop-ups or footnotes. Or how much pop-ups or footnotes result in the biggest impact. These are interesting topics for further research.

6.2 Results for visible or hidden attributes

As credit card companies often want to focus more on the important aspect of a credit card offer, they are interested to know how to avoid displaying too much information on screen. In the previous subsection, it is shown that using pop-ups and footnotes is a good option. Another possibility is to hide product attributes and include these in a 'General terms and conditions' window.

To quantify the impact of hiding attributes on the actual choice, two different models are compared based on external validity, which are measured by hit rates. The two different models, which are described in subsection 5.2.2, are as follows:

1. Main effects model: main effect for two product attributes (balance transfers and late payment fees).

2. Null model: main effects for visualization technique not included.

The main effects are only applied to the secondary product attributes, because hiding the primary attributes is not realistic. Still, an absolute lack of information about the secondary product attributes (balance transfers and late payment fees) of a credit card offer may be seen as suspicious, even if such characteristics are of little importance in the decision. This can lead to the concept being chosen less likely. Another possibility is that the offer is more likely to be chosen, because it is less cluttered.

The average part-worth utilities for all the models are described in appendix 3. The hit rates for the two models are presented in table 6.6. The hit rates are based on input from 555 respondents. Task order (plain CBC exercise or CBC exercise with hidden attributes first) had no significant effect on hold-out performance. This was detected by running two independent estimation processes.

Table 6.6: Hit rates for models with or without hidden product attributes

Model	Model description	Hit rate
Main effects model	Main effect for two attributes	58%
Null model	No visualization technique applied	56%

The hit rates are almost the same. This implies that excluding secondary attributes from the concept visualization has a very minor effect on the consumer's choice. It is especially remarkable that the null model has a high predictive power. This can be explained by the low attribute importance of the secondary product attributes. The attribute importance of each product attribute for each of the two models is displayed in figure 6.1.

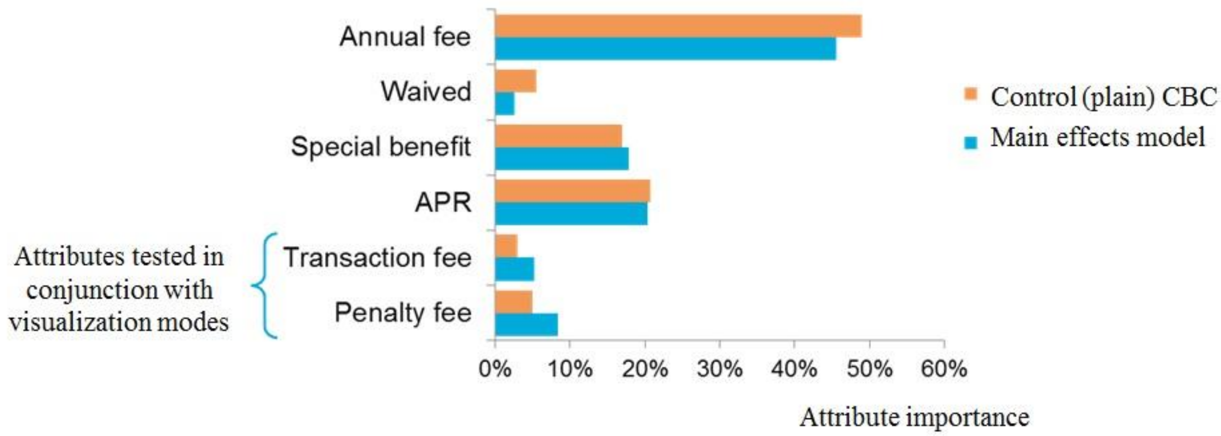


Figure 6.1: Attribute importance when product attributes are hidden

It can be concluded that excluding secondary attributes from the concept visualization has a very minor effect on actual consumer choice. However, there is an exception for which this conclusion may not apply anymore: if one product attribute has an unacceptable attribute level; this is an attribute level that is reason enough to reject a concept (e.g. extremely high late payment fees). The presence of such an unacceptable attribute level would also lead to high(er) importance of the product attribute. Ethically, it is only correct to hide aspects that are of little relevance. Keeping this in mind, this research suggests to keep the information on the main screen short and relevant.

It would have been interesting to know what happens if an unacceptable attribute level is included. Although it may not be ethical, it would be interesting to know what happens if one excludes one of the most important product attributes in the concept visualization. Will the choice for these product concepts be significantly lower, because respondents are very suspicious about these offers, or does this not have any significant impact at all? These are interesting topics for further research.

6.3 Results for font size variations

The last visualization technique that is studied in this thesis is the technique with font size variations. We often see advertisements that play with font sizes to emphasize more favorable aspects, aiming to steer consumer choices based on aspects shown in larger font sizes. it is expected in this study that a product concept is likely to be more attractive if favorable aspects are displayed with a larger font size. For example, showing $0 annual membership fee in a large font size can make the product concept more appealing. Or showing a special benefit in a large font size can make the offer more interesting for the consumer. Annual membership fee, special benefits and annual percentage rate (APR) are the product attributes that are tested with a normal or large font size.

It is also expected that a product concept can be more attractive if less favorable aspects are displayed with a smaller font size. For example, showing late payment fees in a small font can make the product concept more attractive. Balance transfers and late payment fees are the product attributes that are tested with a normal or small font size.

In reality, often more than one attribute level is displayed with a large font size. Therefore, it would be interesting to compare a model with the main effect for all attributes with a model without any visualization techniques. Also, an interaction effect can be included for the most important attribute; waived annual membership fee. This model is interesting to test as well. To quantify the impact of font size variations on the actual choice, the models are compared based on external validity, which are measured by hit rates. In short, the different models are as follows:

1. Full model: main effects for all attributes.

2. Full model with interaction: first-order interaction for the most important attribute (waived annual membership fee).

3. Null model: main effects for visualization technique not included.

The average part-worth utilities for all the models are described in appendix 3. The hit rates for the three models are presented in table 6.7. The results are based on the 284 respondents who first saw the CBC exercise with visualization effects. Respondents who saw the control (plain) CBC exercise first seem to be unaffected by the font size. This can possibly be caused by a learning effect.

Table 6.7: Hit rates for models with or without font size variations

Model	Model description	Hit rate
Full model	Main effects for all attributes	55%
Full model with interaction	First-order interaction for most important attribute	55%
Null model	No visualization technique applied	53%

The hit rates for the full model without interaction and the null model are almost the same. This implies that font size variations have a very minor effect on consumer choice. Although the total impact on actual choice seems small based on the difference in hit rates, a larger font size is an effective driver of preference. This can be seen in figure 6.2, where it is measured how often the bigger font size is preferred over the smaller font size. These results are based on the utilities on a respondent level for the full model with interaction for respondents who saw the CBC exercise with visualization effects first.

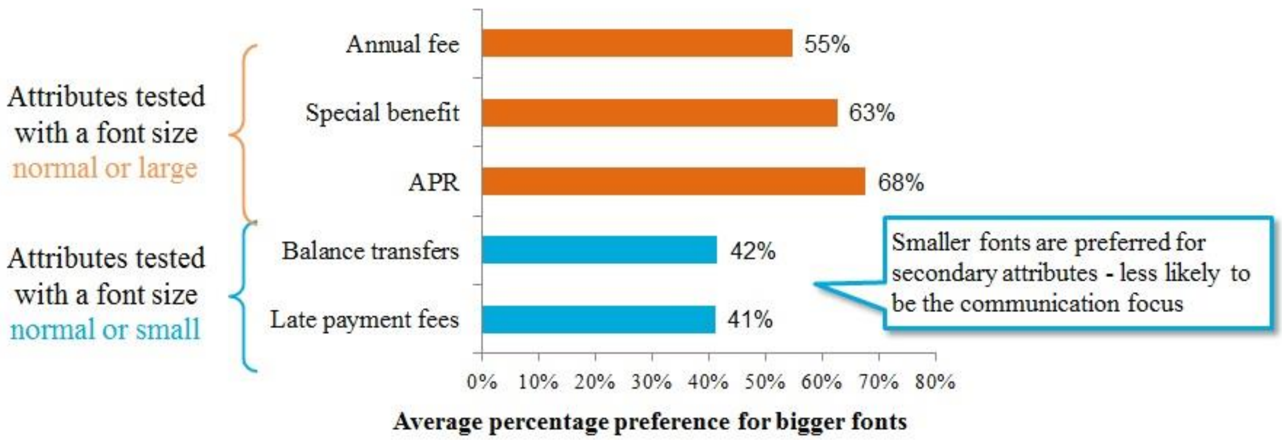


Figure 6.2: Effective drivers of preference

The primary product attributes are preferred with a larger font size and the secondary product attributes are preferred with a smaller font size. So, the font sizes are in line with their relative importance. In order to find out whether the preference for a large font size depends on the level content, the percentage of times that the large font size is preferred for each level of the attribute APR is calculated. Table 6.8 shows that a large font size is almost preferred for each level of the attribute APR, so bigger is better. This also holds for the other product attribute levels. Thus, the preference for a large font size is independent of the level content. This means that it would be easy to lead or even mislead consumer choices by playing with font sizes.

Table 6.8: The relationship between the preference for large font sizes and level content

Attribute level	Percentage of times the larger font size is preferred over the smaller font size
0% Intro APR – 6 months	66%
0% Intro APR – 12 months; after that 10.99 – 17.99%	67%
0% Intro APR – 12 months; after that 12.99 – 22.99 %	65%
0% APR	84%
15% APR	48%
20% APR	76%

This study has quantified the impact of font size variations on actual choice. Large, normal and small font sizes were used. It would have been interesting to know the optimal font size that makes an offer the most attractive, while all information shown on screen is still readable. This is an interesting topic for further research.

7. Discussion

The goal of this thesis is to quantify how the visual attribute representations affect consumers' preferences. The effect of attribute level visualization on consumer choice behavior is explored in the empirical application on credit cards for three conditions: pop-ups and footnotes, hidden attributes and font size variations. It is proven that especially using pop-ups or big font sizes have an effect on consumer choice behavior. However, there are some limitations in the study design that may have impacted the final outcome. These limitations are discussed in this section.

The subject of this thesis is very relevant and valuable for improving marketing studies by increasing realism of the choice options. In the previous section some interesting topics for further research were suggested, which can make the choice options even more realistic. The possibilities for further research are discussed in more detail in this section as well.

Limitations

This thesis used a choice-based conjoint method. This is a very useful tool for marketing analysis. It provides clear insights in what the market looks like and what the potential power of a product is about to become in the near future. However, researchers should be careful when using a choice-based conjoint method. The number of attributes and choice tasks affect the behavior of the respondents of a CBC study. Both too few and too many product attributes can have a negative impact on the results. Also, neglecting barriers and promotion effects are not fully captured by the method. More examples of the limitations of using a CBC method are described in section 2.4. Although there are certain disadvantages, a CBC method can still be used to give accurate estimates of market shares and insights into the preferences for certain product attributes that can be exploited for strategic purposes.

The part-worth utilities are obtained using a MCMC procedure from many different starting points. It should be checked when convergence has been reached. Convergence has been achieved when the draws do not deviate much from the posterior anymore. It is hard to determine upfront when the convergence will be reached. The number of iterations for burn-in already have to be determined, before the MCMC procedure even starts. A possible solution is to use a very high number of iterations to make sure convergence should have been reached. However, this is very time consuming. So, there is a trade-off between the time needed to run the MCMC procedure and the accuracy of the estimates. As technology is improving rapidly, I expect that this would not be an issue anymore in the near future. Market researchers should just select a very high number of iterations for burn-in and it will not take much time before they obtain accurate estimates.

There were some limitations regarding the empirical credit card study design. It was not possible to test the impact of the three visualization techniques on the same respondents. The survey would then have taken too much time for the respondents to finish it. This would have caused the results to be less reliable, as respondents are not likely to give reliable answers anymore when they are bored. Also, the percentage of respondents that do not finish the survey would have been higher. On the other hand, it could be argued that it is better that respondents were only exposed to one CBC task with a visualization technique. Otherwise, a possible learning effect could occur, which can influence the results.

Further research

I have quantified the impact of using pop-ups, footnotes, hiding product attributes and varying font sizes of the choice options on the actual choice. It makes choice options more realistic to the respondents, which in the end leads to more accurate measures of preferences for certain attribute levels. This thesis offers a valuable contribution to the field of marketing research and conjoint methodologies, and there are many possibilities for further research that can further increase the realism of choice options.

Certain topics for further research that are linked to the visualization techniques discussed in this thesis are the effects of combining both pop-ups and footnotes for one product concept on actual choice, compared to only using the same amount of pop-ups or footnotes. Or how much pop-ups or footnotes result in the biggest impact. It can also be examined what happens if one excludes one of the most important product attributes in the concept visualization? Will the choice for these product concepts be lower, because respondents are very suspicious about these offers, or does this not have any significant impact at all? When the effect of using font size variations was studied, only small, normal and big font sizes were used. It would have been interesting to know the optimal font size that makes an offer the most attractive, while keeping all information shown on screen readable. Finally, it can also be taken a step further, by quantifying the impact of a combination of the different visualization techniques applied to choice options on the actual choice.

Other visualization effects that could have an impact on preferences are using a different order of characteristics between product concepts, the tone/length of equivalent attribute levels (e.g. ‘Annual membership fee: $95,00’ versus ‘Every year, you will be charged a fixed amount of $95.00’) and graphical icons versus text and the color of attribute levels (e.g. the color red draws attention and can be used to highlight a price cut). Market researchers can make the choice options even more realistic by using animated three-dimensional product concepts and by designing virtual shopping environments with realistic shelf spaces. Furthermore, it would be useful to study how consumers really perceive exercises using visualization effects. Do they feel they are more realistic and does this help them in making more realistic choices, or is it confusing and/or annoying? And are there any significant differences between market segments?

8. Conclusions

The goal of this thesis was to quantify how the visual attribute representations affect consumers' preferences. The study is of practical relevance, because including visualization results in a research design that better replicates reality, creates more realistic choice exercises, and potentially possesses higher external validity. Also, it allows recommendations to be made on how to communicate the level of an attribute to yield the desired results (within ethical boundaries).

The effect of attribute level visualization on consumer choice behavior was explored for three different conditions. The first visualization technique concerns pop-ups and footnotes that can be included in the product concepts. Instead of presenting attribute levels using plain and descriptive text, a concise and captivating text is used to cover the most important information. The remaining information is given in the pop-up or footnote. The second condition concerns hidden information. If this condition is used, one can quantify the impact of hiding certain attribute levels that are only accessible through a "General terms and conditions" window. The third condition concerns font size variations. Font size variations can be used to emphasize or play down attribute levels or aspects of them.

The study concerns credit cards, because credit cards are important products, everybody knows what the products are about and they are offered by many competitors using various communication styles. As the choice of a credit card has a long-term impact on the buyer's financials, consumers are assumed to look precisely for the best available offer. That is not an easy exercise, because a credit card offer is described by various attributes and not every consumer is completely aware of the consequences of a particular attribute level. Credit cards are offered by many banks. Each bank uses different communication styles; this makes it difficult to directly compare the offers. For the questionnaire, the different offers in the choice options were simplified by having the same order for the product attributes and by using the same font type and overall design for each offer. This makes it a bit easier for consumers to compare the offers, while the impact of the different visualization techniques can still be studied.

The choice behavior was investigated using a choice-based conjoint (CBC) study. Each respondent has decided which of the product concepts shown in several successive choice tasks he preferred. Because people are assumed to react differently to a company's specific credit card offering, the model should be able to deal with heterogeneity. Individual-level "part-worth" utilities were allowed for each respondent. Here, the hierarchical Bayes (HB) estimation for the mixed logit model was used, because of its ability to provide reasonable estimates for the utilities, given only a few choices by each respondent.

It is proven that consumers gravitate to concise and captivating descriptions combined with additional information offered by means of pop-ups or footnotes, instead of having all information readily, but plainly, available on screen. Pop-ups are proven to be more effective than footnotes. Font size variations have an impact as well. For the most important attributes, bigger is better (independently on the content). This also goes for attribute levels that are not very beneficial for the consumers, such as a high annual membership fee. A bigger font size may give the impression that the attribute level must be attractive. The study presents an insight in how easy it is to affect consumer choices and how big the impact can be from both a consumer's and company's perspective.

This thesis offers a valuable contribution to the field of marketing research and conjoint methodologies and there are many possibilities for further research, such as using animated three-dimensional product concepts and by designing virtual shopping environments with realistic shelf spaces, which can increase the realism of choice options even more. I hope this thesis inspires other researchers to conduct more research on the subject of the impact of the visualization of choice options on actual choice, because there are many interesting topics for further research that could lead to improvements for marketing research.

9. Acknowledgements

I would like to acknowledge and offer my sincere gratitude to some of the people who have helped me throughout my studies and especially during the period I was writing this thesis.

At a part of the bachelor program, each student is required to write a bachelor thesis. I was always very intrigued by how people decide what product is best for them. What features convinced them to buy a specific product? And how can this really be tested? This has led to my subject for the bachelor thesis, which was about the effect of ‘best-worst scaling’ applied to new products. I was very happy that I had the opportunity to learn from my supervisor, Dr. A.J. Koning. Together with the co-reader Prof. Dr. P.H.B.F. Franses, they showed me that it can be fun to combine marketing and quantitative methods. This motivated me to enroll for the master program ‘Quantitative marketing’.

I wanted to combine the writing of my master thesis with an internship. During the internship, I wanted to learn how to apply econometric models in practice. I want to thank Prof. Dr. D. Fok. He helped me to search for companies that offered interesting internships. He put me in contact with SKIM, an international market research company specialized in conjoint analysis. Conjoint analysis is all about studying choice behavior/ decision making and this seemed very interesting to me.

At SKIM, I was jointly supervised by Chief Methodology Officer G. Loosschilder, Research Director K. van der Wagt and Project Manager C. Borghi. They are all very experienced in conjoint analysis and Bayesian methods, and they contributed to my understanding of this material. They have offered me both technical and financial support. Also, they have deepened my interest in Bayesian methods. I am most grateful for all the knowledge they have shared with me and it was a pleasure to work with them.

The supervisor from the Erasmus University Rotterdam who was assigned to me is Prof. Dr. P. Goos. He is an expert in Bayesian econometrics and I am extremely fortunate that he helped me during the process of writing my master thesis. His feedback was always very valuable to me. He focused on the main story of the thesis, but also kept an eye for details. The latter helped me to work even more accurate. Also, I would like to thank my co-reader A. Ruseckaite for providing points for improvement. I am delighted to thank both of them for their support.

On the personal side, I want to thank my parents and grandparents. They have always supported me and motivated me to strive for the best results.

10. References

Allenby, G.M., Neeraj, A. & and Ginter, J.L. (1995), 'Incorporating prior knowledge into the analysis of conjoint Studies', *Journal of Marketing Research*, Vol. 32, pp. 152-162.

Allenby, G.M. and Rossi, P.E. (1999), 'Marketing models of consumer heterogeneity', *Journal of Econometrics*, Vol. 89, pp. 57-78.

Bauwens, L., Lubrano, M. & Richard, J. (1999), 'Bayesian inference in dynamic econometric models', *Oxford University Press.*

Bose, R.C. (1942), 'A note on the resolvability of balanced incomplete block designs', *The Indian Journal of Statistics*, Vol. 6(2), pp. 105-110.

Carroll, J.D. and Green, P.E. (1995), 'Psychometric methods in marketing research: part 1, conjoint analysis', *Journal of Marketing Research*, Vol. 32(4), pp. 385-391.

Casella, G. and George, E.I. (1992), 'Explaining the Gibbs sampler', *The American Statistican*, Vol. 46(3), pp. 167-174.

Chib, S. and Greenberg, E. (1995), 'Understanding the Metropolis-Hastings algorithm', *The American Statistican*, Vol. 49(4), pp. 327-335.

Desarbo, W., Ramaswarny, V. & Cohen, S. (1995), 'Market segmentation with Choice-based Conjoint analysis', *Marketing Letters*, Vol. 6(2), pp. 137-147.

DeShazo, J.R. and Fermo, G. (2002), 'Designing choice sets for stated preference methods: the effects of complexity on choice consistency', *Journal of Modelling in Management*, Vol. 5(2), pp. 94-123.

Ding, M., Grewal, R. & Liechty, J. (2005), 'Incentive-aligned conjoint analysis', *Journal of Marketing Research*, Vol. 42(1), pp. 67-82.

Feldman J. K. (2003), 'Use of conjoint analysis for the design of new website', *ISixSigma newsletter.*

Firth, D. (2008), 'Bias reduction of maximum likelihood estimates', *Biometrika*, Vol. 80(1), pp. 27-38.

Gelman, A. and Rubin, D.B. (1992), 'Inference from iterative simulation using multiple sequences', *Statistical Science*, Vol. 7(4), pp. 457-472.

Gelman, A., Roberts, G.O. and Gilks, W.R (1996), 'Efficient Metropolis jumping rules', *Bayesian Statistics 5 – Oxford University Press*, pp. 599-607.

Green, P.E. and Rao, V.R. (1971), 'Conjoint measurement for quantifying judgmental data', *Journal of Marketing Research*, Vol. 8(3), pp. 355-363.

Green, P.E. and Srinivasan, V. (1978), 'Conjoint analysis in consumer research: Issues and outlook', *Journal of Consumer Research*, Vol. 5, pp. 103-123.

Green, P.E. and Srinivasan, V. (1990), 'Conjoint analysis in marketing: new developments with implications for research and practice', *Journal of Marketing*, Vol. 54, pp. 3-19.

Guba E.G. and Lincoln, Y.S. (1994), '*Handbook of qualitative research*', pp. 105-106.

Hauser, J. and Glen, L. (1979), 'Assessment of attribute importance and consumer utility functions, *Journal of Consumer Research*, Vol. 5, pp. 251-262.

Iyengar, S.S. and Lepper, M.R. (2000), 'When choice is de-motivating: can one desire too much of a good thing?', *Journal of Personality and Social Psychology*, Vol. 79, pp. 995-1006.

Jeffreys, H. (1961), 'Theory of Probability', *Clarendon Press, Oxford*, 3rd edition.

Johnson, R.M. (2000), ‘Monotonicity constraints in Choice-based Conjoint with Hierarchical Bayes’, *Sawtooth Software Research paper series.*

Kara, A., Kaynak, E. & Kucukemiroglu, O. (1994), ‘Credit card development strategies for the youth market: the use of conjoint analysis’, *International Journal of Bank Marketing,* Vol. 12(6), pp. 30-36.

Kass, R.E., Carlin, B.P., Gelman, A. & Neal, R.M. (1998), ‘Markov Chain Monte Carlo in practice: a roundtable discussion’, *The American Statistican,* Vol. 52(2), pp. 93-100.

Lancaster, T. (2004), ‘An introduction to modern Bayesian econometrics’, *Blackwell, Malden, MA.*

Lee, J., Cho, Y., Lee, J.D. & Lee, C.Y. (2006), ‘Forecasting future demand for large-screen television sets using conjoint analysis with diffusion model’, *Elsevier – Technological forecasting and Social Change,* Vol. 73(4), pp. 362-376.

Louviere, J.J. and Woodworth, G. (1983), ‘Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data’, *Journal of Marketing Research,* Vol. 20(4), pp. 350-367.

Luce, R. D. and Tukey, J.W. (1964), ‘Simultaneous conjoint measurement – a new type of fundamental measurement’, *Journal of Mathematical Psychology,* Vol. 1, pp. 1-27.

McCulloch, R.E., Polson, N.G. & Rossi, P.E. (2000), ‘A Bayesian analysis of the multinomial probit model with fully identified parameters’, *Journal of Econometrics,* Vol. 99(1), pp. 173-193.

McFadden, D.L. (1974), ‘Conditional logit analysis of qualitative choice behavior’, *Frontiers in Econometrics’,* pp. 105-142.

Meiβner, M., Scholz, S.W. & Decker, R. (2010), ‘Using eye tracking and mouselab to examine how respondents process information in CBC’, *Sawtooth Software Conference 2010,* pp. 151-170

Mueller, S., Lockshin, L. & Louviere, L.L. (2010), ‘What you see may not be what you get: asking consumers what matters may not reflect what they choose’, *Marketing Letters,* Vol. 21(4), pp. 335-350.

Myung R. (2003), ‘Conjoint analysis as a new methodology for Korean typography guideline in web environment’, *International journal of industrial ergonomics*.

Nierop, E. van, Fok, D. & Franses, P.H. (2008), ‘Interaction between shelf layout and marketing effectiveness and its impact on optimizing shelf arrangements’, *Marketing Science,* Vol. 27(6), pp. 1065-1082.

Peral, B.P., Rey, J.R.B. & Ramos, A.F.V. (2012), ‘A study of consumer preferences for e-retailers’ attributes: an application of conjoint analysis’, *International Journal of Management Science and Information Technology*, Vol. 1(3), pp. 37-67.

Pullman, M.E., Dodson, K.J. & Moore, W.L. (1999), ‘A comparison of conjoint methods when there are many attributes’, *Marketing letters,* Vol. 10(2), pp. 125-138.

Rooderkerk, R.P, Bijmolt, T.H.A. & Heerde, H.J. (2011), ‘Incorporating context effects into a choice model’, *Journal of Marketing Research*, Vol. 48(4), pp. 767-780.

Rossi, P.E., Allenby, G.M. & McCulloch, R. (2006), ‘Bayesian statistics and marketing’, *John Wiley & Sons, New York.*

Ryan, M., McIntosh, E. & Shackley, P. (1998), ‘Methodological issues in the application of conjoint analysis in health care’, *Health Economics,* Vol. 7(4), pp. 373-378.

Sandor, Z. and Wedel, M. (2005), 'Heterogeneous conjoint choice designs', *Journal of Marketing Research*, Vol. 42(2), pp. 210-218.

Sonnier, G., Ainslie, A. & Otter, T. (2007), 'Heterogeneity distributions of willingness-to-pay in choice models', *Quantitative Marketing and Economics*, Vol. 5(3), pp. 313-331.

Thurstone, L.L (1927), 'A law comparative judgment', *Psychological review*, Vol. 34, pp. 273-286.

Tierney, L. (1994), 'Markov chains for exploring posterior distributions (with discussion)', *The Annuals of Statistics*, Vol. 22(4), pp. 1701-1762.

Train K. (2001), 'A comparison of Hierarchical Bayes and maximum simulated likelihood for mixed logit', *Working Paper*, Department of Economics, University of California, Berkeley.

Wen, D. and Levy, M.S. (2001), 'BLINEX: a bounded asymmetric loss function with application to Bayesian estimation', *Communication in Statistics – Theory and Methods*, Vol. 30(1), pp. 147-153.

Williams, P. and Kilroy, D. (2000), 'Calibrating price in ACA: The ACA price effect and how to manage it', *Proceedings of the Sawtooth Software Conference*, pp. 81-95.

Wittink, D., Vriens, M. & Burhenne, W. (1994), 'Commercial use of conjoint analysis in Europe: Results and critical reflections', *International Journal of Research in Marketing*, Vol. 11(1), pp. 41-52.

Wonder, N., Wilhelm, W. & Fewings, D. (2008), 'The financial rationality of consumer loan choices: revealed preferences concerning interest rates, down payments, contract length, and rebates', *The Journal of Consumer Affairs*, Vol. 42(2), pp. 243-270.

Zellner, A. (1971), 'The Bayesian method of moments (BMOM): Theory and applications', *Advances in Econometrics*, Vol. 12, pp. 85-105.

A. Appendices

A.1 Questionnaire

S1

What is your age?

…. Years of age

If age < 18 or > 65 → Disqualify

S2

What is your professional situation?

1. Employed, full-time (at least 35+ hours per week)
2. Employed, part-time (between 10-34 hours per week)
3. Student
4. Homemaker
5. Retired
6. Disabled
7. Currently not employed

S3

Do you currently have one or more **credit** cards? (see below for an explanation)

1. Yes
2. No

S4

Do you currently have one or more **debit** cards? (see below for an explanation)

1. Yes
2. No

Credit cards enable you to pay for goods or services by lending money for an interest. Typically you receive a monthly bill stating your expenses at the end of each month.
Debit cards are cards that are linked to your bank account. You cannot use them to pay for goods or services if you don't have enough money in your bank account to cover your payment.

S5 (if S3='2')

How likely is it that you will apply for a credit card in the next 6 months?

1. Extremely likely
2. Rather likely
3. Neither likely nor unlikely → Disqualify
4. Rather unlikely → Disqualify
5. Extremely unlikely → Disqualify

S6 (if S3='1')

How many **credit** cards do you have? …

S7 (if S3='1')

Please think of all transactions you perform on your **credit** card(s).

In an average month, how many **credit card transactions** do you make?
Please **do not consider** the **transactions** you make with your **debit** card(s).

1. Less than once per month → Disqualify
2. 1 to 5 times per month (i.e. about once per week)
3. 6 to 9 times per month
4. 10 to 19 times per month (i.e. about one every two days)
5. 20 to 39 times per month (i.e. about one every day)
6. 40 to 49 times per month
7. 50 to 70 times per month (i.e. about two every day)
8. More than 70 times per month
9. I don't know

S8 (if S3='1')

Please think of the **credit card you use most often**. What type of credit card is it?

1. American Express
2. Visa
3. Mastercard
4. Discover

S9 (if S3='1')
Which bank or credit institute is providing the credit card you use most often?
[randomize list]

1. American Express
2. Bank of America
3. Bank of New York Mellon
4. Barclays Group
5. BB&T
6. Capital One
7. Citibank
8. Discover
9. Fifth Third Bank
10. HSBC
11. ING
12. JPMorgan Chase
13. Northern Trust
14. PNC
15. Royal Bank of Scotland
16. Suntrust
17. Wachovia
18. Wells Fargo
19. Other, please specify:

(this answer: to be used as credit card provider for all credit cards in the CBC)

First conjoint module

Introduction for first CBC pt1:
Please imagine you are in the process of choosing a new credit card from your current bank or credit institute.

On the screen, you will see three credit cards. We would like to ask you to select the one you would **get to use for yourself**. You can assume that your **application** for any credit cards on the screen would **always be accepted**.

[example screen]

[to be repeated at the bottom of every choice task]

If you want to have more information about how credit cards work and the meaning of their features, please click *<u>here</u>* *(http://www.federalreserve.gov/creditcard/flash/offerflash.html)* (Opens in a new window)

First task

This task is a fixed task with 3 products, A B and C. This is to send to each CBC module a same number of people with a preference for A, B and C. Let's say that the preference for the three is (across the sample) 70-20-10. The percentage of people seeing CBC1 that have a preference for A should be 70%, that have a preference for B is 20%, that have a preference for C is 10%. Same for the other exercises. This is to make sure to have the same "base case" shares for all modules

Which of the following credit cards would you like to get for yourself?

Introduction for first CBC pt2:

We will ask you to perform a choice between credit cards 8 times in a row. Please consider each choice you make as a new situation independent from the previous ones.

Extra introduction for CBC1 (*click here to see level*)

In this exercise, the explanation of certain credit card features is shortened to their most relevant aspects. This is to make your choice easier. However, to see the detailed explanation of a given feature, you can click on the symbol:

[example screen with pop-up]

Extra introduction for CBC2 (*click here for full conditions*)

In this exercise, only the most relevant information may be included for some credit cards. This is to make your choice easier.

[example screen]

However, if you wish to know the Penalty Fee conditions for card #2, you can click on the link at the bottom of the credit card explanation. A pop-up containing all relevant information will open on your screen.

Please note that **different credit cards** will have **different conditions for the features that are not shown**.

Furthermore, these conditions are going to be **different** on **each screen**.

After first CBC, questions D1 - D4 are asked

Introduction for the second CBC
We now start the second round of questions. Once again, please select the credit card you would **get to use for yourself**. You can assume that your **application** for any credit cards on the screen would **always be accepted**. You will have to perform a choice on 8 screens.

CBC Question:
Which of the following credit cards would you like to get for yourself?

Break text
Thanks for your answers! We would like to take a break to ask you a few questions about yourself before resuming with the exercise.

D1
How many people 16 years of age or older currently reside in your household, including yourself?

1. 1
2. 2
3. 3
4. 4
5. More than 4

D2
What is your gender?

1. Male
2. Female

D3
What is the highest level of education you have completed?

1. Some high school, no diploma
2. High school graduate or equivalent
3. Some college, no degree
4. Bachelor's or Associate's degree

5. Graduate or professional degree
6. Other, please specify:
7. Prefer not to answer

D4

Which of the following ranges best represents your annual household income before taxes?

1. Under $30,000
2. $30,000 - $39,999
3. $40,000 - $49,999
4. $50,000 - $59,999
5. $60,000 - $69,999
6. $70,000 - $99,999
7. $100,000 - $119,999
8. $120,000 - $149,999
9. $150,000 or more
10. Prefer not to answer

B1 (skip if S3='2')

Have you every failed to make a payment on time or exceeded the credit limit of your credit card?

1. Yes, more than once
2. Yes, only once
3. No, never

B2 (skip if S3='2')

Are you aware of what happens when you fail to make a payment on time or exceed the credit limit of your credit card?

1. Your APR (annual purchase rate) is raised to 29.99% indefinitely
2. Your APR (annual purchase rate) is raised to 29.99% until you make your payments
3. Your APR (annual purchase rate) is raised to 29.99% for a fixed period of time (one month for most cards)
4. Your credit card is cancelled
5. I don't know

B3

Below you will find a series of statements about credit cards made by people who already answered this questionnaire. Please indicate how much you agree with them with a score from 1 (I definitely agree) to 5 (I definitely disagree)

1. I definitely agree
2. I agree up to a point
3. I don't agree or disagree
4. I disagree up to a point
5. I definitely disagree
6. I don't know

On rows:

1. I think credit cards are a scam
2. I love credit cards because they are so convenient
3. Debit cards are so much better than credit cards
4. It's hard to select the right credit card for me
5. When I have to choose a credit card for me, I can easily find all the information I need to make the right choice
6. Some banks/credit institutes offer really great conditions on credit cards
7. Some banks/credit institutes offer really terrible conditions on credit cards
8. The credit card offerings from the largest banks/credit institutes are pretty much equivalent

B4

What is the APR (annual percentage rate) of your main credit card?
If you are unsure, please give us your best estimate
… %

B5

What is the annual membership fee for your main credit card?
If you are unsure, please give us your best estimate
… $/year

B6

We have almost completed the interview. Do you have any specific comments about this questionnaire that you would like to share with us?

1. Yes: …
2. No

A.2 R code manual

In this manual, the algorithm is presented that is used for one of the choice-based conjoint (CBC) studies for this thesis. This algorithm is implemented in the open source language R and thus everybody can replicate the results without having to pay for commercial software. The whole process of obtaining results for an CBC study from the beginning until the end is described. At the beginning, there is a brief description about the basics in R. Next, the empirical application is described, which is used to obtain the results that are described in section 6.3.

A.2.1 Empirical application - Data

The conducted CBC study resulted in available information of 1.665 respondents. Each respondent was exposed to only one of the three possible visualization conditions; one of them is about the varying font sizes. For the purpose of controlling, each respondent answered choice tasks for a plain CBC as well. In the plain CBC, the product concepts in each choice task were fully described without any visualization effects. Information was available from 553 respondents for the font size visualization condition. Each respondent answered eight choice tasks (including one hold out task). For each choice task, the respondent chose the best option out of three alternatives that are presented on screen.

Each product concept consists of a combination of attribute levels corresponding to five different attributes. All the attributes and levels are displayed in table A.2.1. The "B" means that a large font size is used, and a "S" means a relatively smaller font size is used. In the first column, the membership fee is given. The levels of this attribute can be waived or not. The amount you have to pay for the fees is given by the numbers. If you have to pay for your membership, so it is not 0, then the level of this attribute can be waived ("W") or not waived ("N"). The second column represents the special benefits; in case they are present, you have three different types of benefits. Attribute 3 represents the annual percentage rate (APR), of which there are six different options. B6 represents 'intro APR for 6 months' and after that period a certain percentage, B12h represents 'intro APR for 12 months' and after that period a higher percentage, while B12l is the same but with the same percentage as the one for 6 months. Furthermore, there are also basic APR's without special percentages in the introduction period. The secondary attributes are balance transfers and payment fees. These attributes are presented in the last two columns.

Table A.2.1: Attributes and levels for application of font size visualization

Attribute 1 *Membership fee*	**Attribute 2** *Special benefits*	**Attribute 3** *APR*	**Attribute 4** *Balance transfers*	**Attribute 5** *Payment fee*
0B	BNo	B6	B5	B1
0S	B1%	B12h	B10	B2
199BW	B200	B12l	S5	B3
199BN	BAirmiles	B20	S10	S1
199SW	SNo	B15		S2
199SN	S1%	B0		S3
67BW	S200	S6		
67BN	SAirmiles	S12h		
67SW		S12l		
67SN		S20		
40BW		S15		
40BN		S0		
40SW				
40SN				

<u>*Constraints*</u>

If you want to obtain accurate and useful estimates for the utilities for all the attribute levels for all the respondents, you might want to consider using constraints. In this case, the following constraints are used:

Attribute 1

0B > 40BW > 67BW > 199BW
0S > 40SW > 67SW > 199SW
0B > 40BN > 67BN > 199BN
0S > 40SN > 67SN > 199SN
40BW > 40BN
40SW > 40SN
67BW > 67BN
67SW > 67SN
199BW > 199BN
199SW > 199SN

Attribute 2

B1% > BNo

S1% > SNo
B200 > BNo
S200 > SNo
BAirmiles > BNo
SAirmiles > SNo

Attribute 3

B12l > B6
B12l > B12h
S12l > S6
S12l > S12h
B0 > B15 > B20
S0 > S15 > S20

Attribute 4

B5 > B10
S5 > S10

No constraints are applied to the last attribute.

A.2.2 Model in R

A MCMC algorithm is used to estimate a hierarchical mixed logit model. The algorithm uses a hybrid Gibbs sampler, which was explained in detail in section 4.3. The dependent variable is discrete. The independent variables are discrete as well with optional order constraints.

The R-code requires the package 'ChoiceModelR'. This package includes the function 'choicemodelr' that implements the required MCMC algorithm. The basic structure of the code is derived from the package 'bayesm'. Significant modifications were made to allow constraints on estimated parameters and to reduce the run time. The function choicemodelr consists of five arguments: data, xcoding, mcmc, constraints and options. Information on each argument is given below.

Argument 1: data

The argument 'data' represents a data frame. The column variables of the data frame that are used are as follows: ID,"Set","Alt","X1","X2","X3","X4","X5","Y". If you want to use input from Excel, you have to save your input in a CSV-file and you should contain headers.

The first column represents the number of the respondent. The second column contains the choice task number for this respondent. The third column contains the alternative number within this choice task. The independent variables are represented in the columns X1 until X5. The numbers are matched to specific attribute levels. The dependent variable is contained in the last column. Only in the first row of the choice task data a non-zero value is presented; it takes a integer value in the range [1, 3] (because there are three alternatives). For example, in figure A.2.1, the first input that is used for the application is provided. The rows for the first two choice tasks for the first respondent are presented. One can observe that in the first choice task, he chose alternative 3, because the corresponding value in the column 'Y' is 3. The second non-zero value is 1; this means that he chose alternative 1 in the second choice task.

```
ID,"Set","Alt","X1","X2","X3","X4","X5","Y"
1,1,1,3,1,8,1,5,3
1,1,2,14,2,10,4,3,0
1,1,3,9,6,6,3,1,0
1,2,1,2,5,3,2,4,1
1,2,2,6,4,4,4,2,0
1,2,3,9,7,5,1,1,0
```

Figure A.2.1: Example used to describe set-up input data

Argument 2: xcoding

The vector 'xcoding' specifies the way in which each attribute is coded; 0 = categorical and 1 = continuous. In this case, there are five discrete/categorical independent variables; thus, a vector with five zeros is used as input for this argument.

Argument 3: mcmc

The argument 'mcmc' is a list with three arguments; list(R, use, s). 'R' represents the total number of iterations of the MCMC chain to be performed. It is recommended to choose R = 20,000. The argument 'use' determines the number of iterations that is used in parameter estimation; 10,000 is recommended. In that case, the first 10,000 iterations are used as a burn-in sample. The argument 's' is a scaling parameter, which is used to adjust the standard deviation of random draws of the parameters during the random walk metropolis step of the MCMC chain. By default, it is set at s = 0.1; to keep acceptance at approximately 30%.

Argument 4: constraints

The constraints are described by matrices containing the values 0, 1 and -1. If one wants to specify constraints, a constraint matrix is required for every attribute. In case you do not have any constraints for specific attributes, you should create a zero-matrix for these attributes. Square matrices are needed, because the dimensions must be equal to the number of levels of the attribute it represents. The following equations are needed:

C[i,j] = 1 if beta(i) > beta(j);
C[i,j] = -1 if beta(i) < beta(j);
C[i,j] = 0 otherwise.

The lower-triangular and diagonal portions of the matrix have no added value; thus, the values in these positions can be ignored. According to the constraints that are described above, the code for the constraints that is described in figure A.2.2a-e should be used:

```
constraints = list(c1,c2,c3,c4,c5)

c1 = matrix(c(0,0,1,1,0,0,1,1,0,0,1,1,0,0,
0,0,0,0,1,1,0,0,1,1,0,0,1,1,
0,0,0,1,0,0,-1,0,0,0,-1,0,0,0,
0,0,0,0,0,0,0,-1,0,0,0,-1,0,0,
0,0,0,0,0,1,0,0,-1,0,0,0,-1,0,
0,0,0,0,0,0,0,0,0,-1,0,0,0,-1,
0,0,0,0,0,0,0,1,0,0,-1,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,-1,0,0,
0,0,0,0,0,0,0,0,0,1,0,0,-1,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,-1,
0,0,0,0,0,0,0,0,0,0,0,1,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0), ncol=14, byrow=TRUE)

c2 = matrix(c(0,-1,-1,-1,0,0,0,0,
0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,
0,0,0,0,0,-1,-1,-1,
0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0), ncol=8, byrow=TRUE)

c3 = matrix(c(0,0,-1,0,0,0,0,0,0,0,0,0,
0,0,-1,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,-1,-1,0,0,0,0,0,0,
0,0,0,0,0,-1,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,
```

```
0,0,0,0,0,0,0,0,-1,0,0,0,
0,0,0,0,0,0,0,0,-1,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,-1,-1,
0,0,0,0,0,0,0,0,0,0,0,-1,
0,0,0,0,0,0,0,0,0,0,0,0), ncol=12, byrow=TRUE)

c4 = matrix(c(0,1,0,0,
0,0,0,0,
0,0,0,1,
0,0,0,0), ncol=4, byrow=TRUE)

c5 = matrix(c(0,0,0,0,0,0,
0,0,0,0,0,0,
0,0,0,0,0,0,
0,0,0,0,0,0,
0,0,0,0,0,0,
0,0,0,0,0,0), ncol=6, byrow=TRUE)
```

Figure A.2.2a-e: Necessary input for the constraints

Argument 5: options
The argument ‘options’ is a list with three arguments; list(none, save, keep). The argument ‘none’ is set to ‘TRUE’ to estimate a none parameter; the data does not include a row for this. The default option is ‘FALSE’. The second argument is set to ‘TRUE’ if you want to save draws of several estimates. The argument “keep” is the thinning parameter. It defines the number of random draws to save (default is 10).

A.2.3 R-code

In this subsection, the script code to run the analysis using R is described.

```
#René Edelenbosch
# Rotterdam - Erasmus University Rotterdam

#This code is based on the work of Ryan Sermas and John Colias.

#CHOICE MODELING IN R
#-------------------------------------------------------------------------------------------------------
#
```

```
#-------------------------------------------------------------------------------------------------------------

##DETERMINING THE DIRECTORY
getwd()
setwd("C:/Users/rene7687/Desktop/Programming in R")
#Make sure that everything, including the data files, are located in the right directory.

##LOAD REQUIRED LIBRARIES
library(ChoiceModelR)
library(Matrix)

##LOAD DATA FROM CSV-file.
datar <- read.csv("Data_Fixed.csv")
dim(datar)
head(datar)

##MODEL
#-------------------------------------------------------------------------------------------------------------
#choicemodelr is used to estimate the paramaters of the choice model.
#For the determination of convergence of the MCMC chain, set R = 20000 and use = 10000.
#Only discrete independent variables are used.
#The thinning parameter (keep) is set to 10.
#-------------------------------------------------------------------------------------------------------------

xcoding = c(0,0,0,0,0)

mcmc = list(R = 20000, use = 10000, s = 0.1)

options = list(none=FALSE, save=TRUE, keep=10)
##CONSTRAINTS
#-------------------------------------------------------------------------------------------------------------
#One can make a constrained model as well.
#This can be done by using 'constraints'.
#In case you want to turn back to the unconstrained model,
#just delete constraints = constraints in the function out = choicemodelr(...).
#-------------------------------------------------------------------------------------------------------------

constraints = list(c1,c2,c3,c4,c5)

c1 = matrix(c(0,0,1,1,0,0,1,1,0,0,1,1,0,0,
```

```
0,0,0,0,1,1,0,0,1,1,0,0,1,1,
0,0,0,1,0,0,-1,0,0,0,-1,0,0,0,
0,0,0,0,0,0,0,-1,0,0,0,-1,0,0,
0,0,0,0,0,1,0,0,-1,0,0,0,-1,0,
0,0,0,0,0,0,0,0,0,-1,0,0,0,-1,
0,0,0,0,0,0,0,1,0,0,-1,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,-1,0,0,
0,0,0,0,0,0,0,0,0,1,0,0,-1,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,-1,
0,0,0,0,0,0,0,0,0,0,0,1,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0), ncol=14, byrow=TRUE)

c2 = matrix(c(0,-1,-1,-1,0,0,0,0,
0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,
0,0,0,0,0,-1,-1,-1,
0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0), ncol=8, byrow=TRUE)

c3 = matrix(c(0,0,-1,0,0,0,0,0,0,0,0,0,
0,0,-1,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,-1,-1,0,0,0,0,0,0,
0,0,0,0,0,-1,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,-1,0,0,0,
0,0,0,0,0,0,0,0,-1,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,-1,-1,
0,0,0,0,0,0,0,0,0,0,0,-1,
0,0,0,0,0,0,0,0,0,0,0,0), ncol=12, byrow=TRUE)

c4 = matrix(c(0,1,0,0,
0,0,0,0,
0,0,0,1,
0,0,0,0), ncol=4, byrow=TRUE)
```

```
c5 = matrix(c(0,0,0,0,0,0,
0,0,0,0,0,0,
0,0,0,0,0,0,
0,0,0,0,0,0,
0,0,0,0,0,0,
0,0,0,0,0,0), ncol=6, byrow=TRUE)

##CHARACTERIZING FUNCTION OF THE MODEL
out = choicemodelr(datar, xcoding, mcmc = mcmc, options = options, constraints = constraints)
```

A.2.4 Output

Five different statistics are written to the screen after each 100 iterations. There is a similarity to those that can be obtained by using Sawtooth Software's CBC/HB.

Acceptance	Percentage of MCMC draws that are accepted in the Metropolis Hastings step
RLH	The m −root of the likelihood, where m is the average number of alternatives within a choice tasks.
Percent Certainty	The percent difference between the obtained log likelihood and the log likelihood of a chance model.
Average Variance	The average variance of the final estimates of the model coefficients among all respondents.
RMS	The root mean squared of the final estimates of the model coefficients among all respondents.

The estimates of mu and the means of the model coefficients from the distribution of heterogeneity are displayed in a figure (see figure A.2.3).

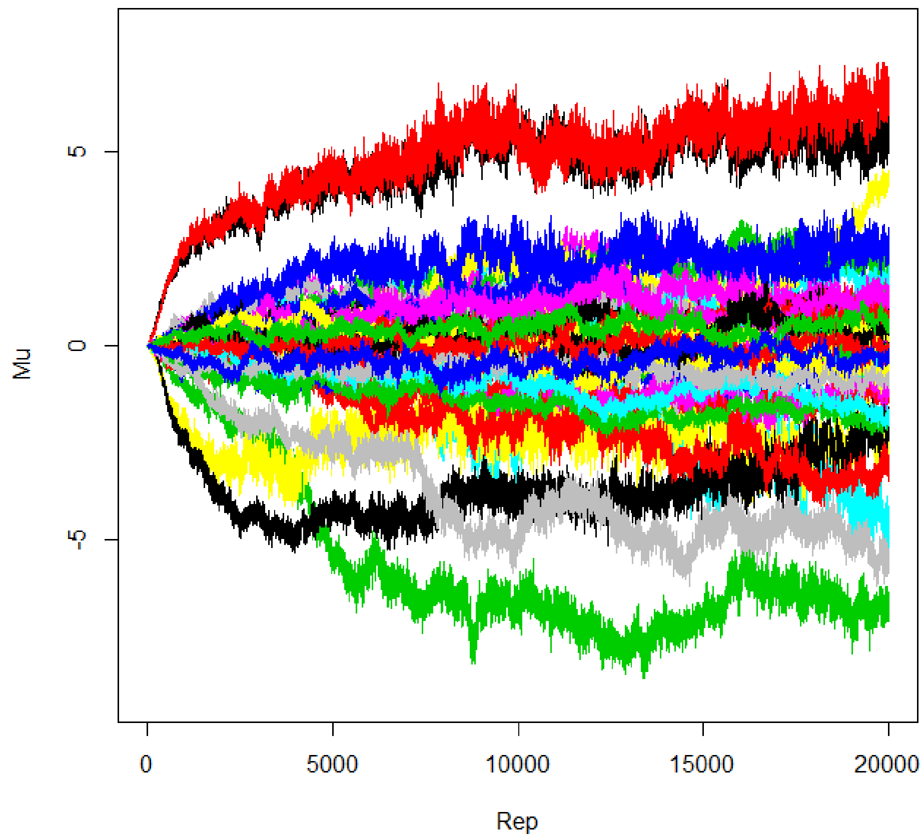


Figure A.2.3: Estimates of mu

At the end of the model estimation, the averages of the MCMC draws of the unit-level model coefficients are written to 'Xbetas.csv' (see table A.2.2). The documentation of the run-time output is written in 'Rlog.txt' (see figure A.2.4). The latest MCMC draws are written to the file 'restart.txt'.

Table A.2.2: A part of the matrix of beta draws

ID	A1B1	A1B2	A1B3	A1B4	A1B5	A1B6	A1B7
1	3,702632	2,459892	-1,87891	-2,23576	-3,16478	-3,6102	1,47374
2	-0,00328	1,400268	-0,5427	-0,55891	-0,16805	-1,3916	-0,26531
3	4,48331	5,427866	-3,31923	-3,37505	-2,48668	-2,57952	-2,5769
4	16,29937	16,46604	-8,02445	-8,07526	-6,06882	-6,10926	-5,68331
5	13,04262	11,67089	-5,59172	-5,68392	-4,02494	-4,12001	-3,74245
6	14,70029	15,73977	-8,22758	-8,28698	-6,04378	-6,09598	-4,21798

Logit Data

==

Attribute	Type	Levels
Attribute 1	Part Worth	14
Attribute 2	Part Worth	8
Attribute 3	Part Worth	12
Attribute 4	Part Worth	4
Attribute 5	Part Worth	6

39 parameters to be estimated.

553 total units.
Average of 3 alternatives in each of 7 sets per unit.
3871 tasks in total.

Table of choice data pooled across units:

Choice	Count	Pct.
1	1353	34.95%
2	1289	33.3%
3	1229	31.75%

MCMC Inference for hierarchical Logit

==

Total Iterations: 20000
Draws used in estimation: 10000
Units: 553
Parameters per unit: 39
Constraints in effect.
Draws are to be saved.
Prior degrees of freedom: 5
Prior variance: 2

MCMC Iteration Beginning...

Iteration	Acceptance	RLH	Pct. Cert.	Avg. Var.	RMS	Time to End
100	0.340	0.404	0.165	0.23	0.25	13:22
200	0.304	0.479	0.319	0.41	0.49	12:54
300	0.301	0.540	0.432	0.63	0.69	12:43

…

9900	0.303	0.832	0.833	13.52	4.34	6:33
10000	0.305	0.834	0.835	14.21	4.40	6:29
10100	0.301	0.831	0.832	14.38	4.39	6:26
			…			
19800	0.302	0.831	0.831	13.17	4.38	0:08
19900	0.301	0.834	0.835	13.43	4.42	0:04
20000	0.304	0.835	0.836	13.79	4.47	0:00

Total Time Elapsed: 12:58

Figure A.2.4: Run-time output

A.3 Results – average part-worth utilities

The average part-worth utilities for all models, that are described in section 6, are presented in this appendix. The average utilities for different models to study the impact of one visualization technique are presented in one table. This makes it easier for comparison. In order to do this, the attribute levels are recoded in small concise descriptions first. The recoded attribute levels can be found in tables A.3.1 – A.3.5. In table 3.1, one can find the coding for the basic attribute levels. In tables A.3.2 – A.3.5, one can find the coding for the additional attribute levels used when a visualization technique is applied. The part-worth utilities are presented in tables A.3.6 – A.3.8.

Table A.3.1: Attribute levels - basic

Attribute	Level	Coding
Annual membership fee	$0.00	A1L1
	$40.00	A1L2
	$67.50	A1L3
	$95.00	A1L4
Waived fee	No annual membership fee for six months, then $x	A2L1
	$x annual membership fee	A2L2
Special benefit	No special benefit	A3L1
	Special benefit: Unlimited 1% Cash Back	A3L2
	Special benefit: $200 Cash Back after you spend $500 in 3 months	A3L3
	Special benefit: Earn 2 Air Miles for each $1 spent	A3L4
Annual percentage rate (APR)	0% Intro APR for 6 months on purchases; after that a variable APR of 10.99% - 17.99% depending on your credit worth	A4L1
	0% Intro APR for 12 months on purchases; after that a variable APR of 10.99% - 17.99% depending on your credit worth	A4L2
	0% Intro APR for 12 months on purchases; after that a variable APR of 12.99% - 22.99% depending on your credit worth	A4L3
	0% APR	A4L4
	15% APR	A4L5
	20% APR	A4L6
Balance transfers/cash advance fees	Balance Transfers: Either $5 or 3% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 3% of the amount of each transaction whichever is greater.	A5L1
	Balance Transfers: Either $10 or 5% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 4% of the amount of each transaction whichever is greater.	A5L2
Late payment fees	Late Payment: Up to $15 if the balance is less than $150; Up to $45 otherwise. Return Payment: Up to $35	A6L1
	Late Payment: Up to $35. Return Payment: Up to $25.	A6L2
	Late Payment: Up to $29 if the balance is less than $1000; Up to $39 otherwise. Return Payment: Up to $23.	A6L3

Table A.3.2: Attribute levels – pop-ups and footnotes

Attribute	Level	Coding
Annual membership fee	Annual membership fee with footnote	A7L1
	Annual membership fee with pop-up	A7L2
	Annual membership fee without visualization	A7L3
Annual percentage rate (APR)	APR with footnote	A8L1
	APR with pop-up	A8L2
	APR without visualization	A8L3
Balance transfers/cash advance fees	Balance transfers with footnote	A9L1
	Balance transfers with pop-up	A9L2
	Balance transfers without visualization	A9L3
Late payment fees	Late payment fees with footnote	A10L1
	Late payment fees with pop-up	A10L2
	Late payment fees without visualization	A10L3
Waived annual membership fee	$0.00 fee not waived	A11L1
	$95.00 fee waived with footnote	A11L2
	$95.00 fee waived with pop-up	A11L3
	$95.00 fee waived without visualization	A11L4
	$95.00 fee not waived	A11L5
	$67.50 fee waived with footnote	A11L6
	$67.50 fee waived with pop-up	A11L7
	$67.50 fee waived without visualization	A11L8
	$67.50 fee not waived	A11L9
	$40.00 fee waived with footnote	A11L10
	$40.00 fee waived with pop-up	A11L11
	$40.00 fee waived without visualization	A11L12
	$40.00 fee not waived	A11L13

Table A.3.3: Attribute levels – visible or hidden attributes

Attribute	Level	Coding
Balance transfers/cash advance fees	Hidden - Balance Transfers: Either $5 or 3% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 3% of the amount of each transaction whichever is greater.	A12L1
	Visible - Balance Transfers: Either $5 or 3% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 3% of the amount of each transaction whichever is greater.	A12L2
	Hidden - Balance Transfers: Either $10 or 5% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 4% of the amount of each transaction whichever is greater.	A12L3
	Visible - Balance Transfers: Either $10 or 5% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 4% of the amount of each transaction whichever is greater.	A12L4
Late payment fees	Hidden - Late Payment: Up to $15 if the balance is less than $150; Up to $45 otherwise. Return Payment: Up to $35	A13L1
	Visible - Late Payment: Up to $15 if the balance is less than $150; Up to $45 otherwise. Return Payment: Up to $35	A13L2
	Hidden - Late Payment: Up to $35. Return Payment: Up to $25.	A13L3
	Visible - Late Payment: Up to $35. Return Payment: Up to $25.	A13L4
	Hidden - Late Payment: Up to $29 if the balance is less than $1000; Up to $39 otherwise. Return Payment: Up to $23.	A13L5
	Visible - Late Payment: Up to $29 if the balance is less than $1000; Up to $39 otherwise. Return Payment: Up to $23.	A13L6

Table A.3.4a: Attribute levels – font size variations (part 1)

Attribute	Level	Coding
Annual membership fee	Big font - $0.00	A14L1
	Big font - $40.00	A14L2
	Big font - $67.50	A14L3
	Big font - $95.00	A14L4
	Normal font - $0.00	A14L5
	Normal font - $40.00	A14L6
	Normal font - $67.50	A14L7
	Normal font - $95.00	A14L8
Waived fee	Big font - No annual membership fee for six months, then $x	A15L1
	Big font - $x annual membership fee	A15L2
	Normal font - No annual membership fee for six months, then $x	A15L3
	Normal font - $x annual membership fee	A15L4
Special benefit	Big font - No special benefit	A16L1
	Big font - Special benefit: Unlimited 1% Cash Back	A16L2
	Big font - Special benefit: $200 Cash Back after you spend $500 in 3 months	A16L3
	Big font - Special benefit: Earn 2 Air Miles for each $1 spent	A16L4

Table A.3.4b: Attribute levels – font size variations (part 2)

Attribute	Level	Coding
	Normal font - No special benefit	A16L5
	Normal font - Special benefit: Unlimited 1% Cash Back	A16L6
	Normal font - Special benefit: $200 Cash Back after you spend $500 in 3 months	A16L7
	Normal font - Special benefit: Earn 2 Air Miles for each $1 spent	A16L8
Annual percentage rate (APR)	Big font - 0% Intro APR for 6 months on purchases; after that a variable APR of 10.99% - 17.99% depending on your credit worth	A17L1
	Big font - 0% Intro APR for 12 months on purchases; after that a variable APR of 10.99% - 17.99% depending on your credit worth	A17L2
	Big font - 0% Intro APR for 12 months on purchases; after that a variable APR of 12.99% - 22.99% depending on your credit worth	A17L3
	Big font - 0% APR	A17L4
	Big font - 15% APR	A17L5
	Big font - 20% APR	A17L6
	Normal font - 0% Intro APR for 6 months on purchases; after that a variable APR of 10.99% - 17.99% depending on your credit worth	A17L7
	Normal font - 0% Intro APR for 12 months on purchases; after that a variable APR of 10.99% - 17.99% depending on your credit worth	A17L8
	Normal font - 0% Intro APR for 12 months on purchases; after that a variable APR of 12.99% - 22.99% depending on your credit worth	A17L9
	Normal font - 0% APR	A17L10
	Normal font - 15% APR	A17L11
	Normal font - 20% APR	A17L12
Balance transfers/cash advance fees	Normal font - Balance Transfers: Either $5 or 3% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 3% of the amount of each transaction whichever is great.	A18L1
	Normal font - Balance Transfers: Either $10 or 5% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 4% of the amount of each transaction whichever is great	A18L2
	Small font - Balance Transfers: Either $5 or 3% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 3% of the amount of each transaction whichever is great.	A18L3
	Small font - Balance Transfers: Either $10 or 5% of the amount of each transfer whichever is greater. Cash Advances: Either $10 or 4% of the amount of each transaction whichever is great	A18L4
Late payment fees	Normal font - Late Payment: Up to $15 if the balance is less than $150; Up to $45 otherwise. Return Payment: Up to $35	A19L1
	Normal font - Late Payment: Up to $35. Return Payment: Up to $25.	A19L2
	Normal font - Late Payment: Up to $29 if the balance is less than $1000; Up to $39 otherwise. Return Payment: Up to $23.	A19L3
	Small font - Late Payment: Up to $15 if the balance is less than $150; Up to $45 otherwise. Return Payment: Up to $35	A19L4
	Small font - Late Payment: Up to $35. Return Payment: Up to $25.	A19L5
	Small font - Late Payment: Up to $29 if the balance is less than $1000; Up to $39 otherwise. Return Payment: Up to $23.	A19L6

Table A.3.5: Attribute levels – font size variations (part 3)

Attribute	Level	Coding
Waived annual membership fee	Big font - $0.00	A20L1
	Small font - $0.00	A20L2
	Big font - $95.00 waived	A20L3
	Big font - $95.00 not waived	A20L4
	Small font - $95.00 waived	A20L5
	Small font - $95.00 not waived	A20L6
	Big font - $67.50 waived	A20L7
	Big font - $67.50 not waived	A20L8
	Small font - $67.50 waived	A20L9
	Small font - $67.50 not waived	A20L10
	Big font - $40.00 waived	A20L11
	Big font - $40.00 not waived	A20L12
	Small font - $40.00 waived	A20L13
	Small font - $40.00 not waived	A20L14

Table A.3.6: Average part-worth utilities - pop-ups and footnotes

Attribute level	Average part-worths full model	Average part-worths single effect model	Average part-worths single effect model + interaction	Average part-worths null model	Attribute level	Average part-worths full model	Average part-worths single effect model	Average part-worths single effect model + interaction
A1L1	4,74	3,01		2,20	**A7L1**	0,86	0,55	
A1L2	0,05	-0,01		-0,14	**A7L2**	3,11	1,98	
A1L3	-1,51	-1,02		-0,66	**A7L3**	-3,97	-2,53	
A1L4	-3,27	-1,97		-1,41	**A8L1**	0,03		
A2L1	2,29	1,47		1,25	**A8L2**	0,09		
A2L2	-2,29	-1,47		-1,25	**A8L3**	-0,13		
A3L1	-2,23	-1,52	-1,42	-1,13	**A9L1**	-0,27		
A3L2	1,77	1,14	1,08	0,77	**A9L2**	0,22		
A3L3	1,59	1,01	1,01	0,77	**A9L3**	0,05		
A3L4	-1,12	-0,63	-0,67	-0,41	**A10L1**	0,01		
A4L1	0,17	0,16	0,03	0,21	**A10L2**	0,19		
A4L2	0,53	0,33	0,45	0,21	**A10L3**	-0,20		
A4L3	0,03	0,09	-0,02	0,06	**A11L1**			3,00
A4L4	1,41	0,88	1,07	0,69	**A11L2**			0,54
A4L5	-0,14	-0,11	-0,07	-0,12	**A11L3**			2,39
A4L6	-2,00	-1,35	-1,47	-1,05	**A11L4**			-4,17
A5L1	0,02	0,08	0,12	0,08	**A11L5**			-5,29
A5L2	-0,02	-0,08	-0,12	-0,08	**A11L6**			1,45
A6L1	0,26	0,20	0,22	0,13	**A11L7**			2,88
A6L2	-0,03	-0,04	-0,04	0,00	**A11L8**			-1,88
A6L3	-0,22	-0,16	-0,17	-0,13	**A11L9**			-2,96
					A11L10			1,97
					A11L11			2,63
					A11L12			-0,10
					A11L13			-0,45

Table A.3.7: Average part-worth utilities – visible or hidden attributes

Attribute level	Average part-worth utilities main effects model	Average part-worth utilities null model	Attribute level	Average part-worth utilities main effects model
A1L1	6,42	4,86	**A12L1**	-0,19
A1L2	1,42	1,08	**A12L2**	0,20
A1L3	-2,23	-1,59	**A12L3**	-0,13
A1L4	-5,61	-4,35	**A12L4**	0,12
A2L1	0,34	0,21	**A13L1**	0,13
A2L2	-0,34	-0,21	**A13L2**	0,14
A3L1	-2,13	-1,68	**A13L3**	-0,20
A3L2	1,12	0,83	**A13L4**	0,01
A3L3	1,49	1,17	**A13L5**	-0,14
A3L4	-0,48	-0,33	**A13L6**	0,06
A4L1	-0,18	-0,14		
A4L2	0,75	0,58		
A4L3	-0,19	-0,11		
A4L4	1,64	1,22		
A4L5	-0,11	-0,01		
A4L6	-1,91	-1,54		
A5L1		0,02		
A5L2		-0,02		
A6L1		0,08		
A6L2		-0,04		
A6L3		-0,04		

Table A.3.8: Average part-worth utilities – font size variations

Attribute level	Average part-worth utilities null model	Attribute level	Average part-worth utilities full model	Average part-worth utilities full model with interaction	Attribute level	Average part-worth utilities full model	Average part-worth utilities full model with interaction	Attribute level	Average part-worth utilities full model with interaction
A1L1	4,06	**A14L1**	7,23		**A17L1**	-1,07	-1,11	**A20L1**	8,07
A1L2	1,24	**A14L2**	2,28		**A17L2**	0,56	0,74	**A20L2**	8,46
A1L3	-0,98	**A14L3**	-2,37		**A17L3**	-0,17	0,11	**A20L3**	-3,32
A1L4	-4,32	**A14L4**	-5,64		**A17L4**	3,29	3,43	**A20L4**	-5,09
A2L1	0,30	**A14L5**	7,43		**A17L5**	-0,69	0,16	**A20L5**	-6,45
A2L2	-0,30	**A14L6**	2,48		**A17L6**	-2,01	-1,71	**A20L6**	-6,80
A3L1	-1,09	**A14L7**	-3,25		**A17L7**	-0,11	-0,88	**A20L7**	-0,99
A3L2	0,66	**A14L8**	-8,16		**A17L8**	1,70	1,10	**A20L8**	-2,53
A3L3	0,81	**A15L1**	0,71		**A17L9**	0,60	-0,11	**A20L9**	0,48
A3L4	-0,38	**A15L2**	-0,47		**A17L10**	1,88	1,99	**A20L10**	-2,37
A4L1	-0,19	**A15L3**	0,30		**A17L11**	-0,49	-0,01	**A20L11**	3,43
A4L2	0,47	**A15L4**	-0,54		**A17L12**	-3,51	-3,72	**A20L12**	1,63
A4L3	-0,01	**A16L1**	-2,02	-2,05	**A18L1**	0,08	0,06	**A20L13**	2,82
A4L4	1,18	**A16L2**	2,32	2,09	**A18L2**	-0,77	-0,57	**A20L14**	2,65
A4L5	0,23	**A16L3**	2,21	2,01	**A18L3**	0,63	0,67		
A4L6	-1,68	**A16L4**	-1,50	-0,80	**A18L4**	0,06	-0,16		
A5L1	0,11	**A16L5**	-1,65	-1,71	**A19L1**	0,47	0,28		
A5L2	-0,11	**A16L6**	0,62	0,85	**A19L2**	-0,75	-1,00		
A6L1	0,16	**A16L7**	0,55	0,49	**A19L3**	0,28	0,29		
A6L2	-0,10	**A16L8**	-0,53	-0,88	**A19L4**	-0,24	-0,13		
A6L3	-0,05				**A19L5**	0,41	0,60		
					A19L6	-0,18	-0,05		