Carlo Borghi

Discrete choice models for marketing

New methodologies for optional features and bundles

Master thesis, defended on November 12, 2009

Thesis advisor: Richard D. Gill

Mastertrack: Mathematics and Science Based Business

Mathematisch Instituut, Universiteit Leiden

Contents

1 Quantitative techniques for marketing **1**
1.1 Conjoint methodologies 2

2 Choice-based Conjoint **4**
2.1 What is CBC conjoint 4
2.1.1 Phases of a CBC conjoint study 5
2.1.2 Different estimation procedures 7
2.1.3 A Study example 8
2.1.4 Scenario simulation methods 11
2.1.5 Limitations of CBC 12
2.2 Discrete choice models 15
2.2.1 Derivation of choice probabilities 15
2.2.2 Utilities and additive constants 16
2.2.3 Utility scale 17
2.3 Logit Model 17
2.3.1 Choice probabilities 17
2.3.2 Estimation procedure 20
2.3.3 Choice among a subset of alternatives 23

3 The Hierarchical Logit model **25**
3.1 Introduction 25
3.2 Hierarchical models for marketing 25
3.3 Inference for hierarchical models 28
3.4 The Hierarchical Bayes multinomial logit model 28
3.4.1 Estimation for the Hierarchical logit model 31

4 Optional features **34**
4.1 Business perspective 34
4.2 Summary of business questions 37
4.3 Scope of our analysis 37
4.4 Methodology 38
4.5 Questionnaire structure 38
4.5.1 Intended result 39
4.6 Assumptions 39

4.7 Simulations . . . 40
4.7.1 Answer generation procedures . . . 41
4.7.2 Set of utilities . . . 41
4.8 Estimation procedures . . . 42
4.9 Measures of fit . . . 45
4.10 Simulation results . . . 45
4.11 Overview of results . . . 48
4.12 Conclusions . . . 51

5 Optional features study **52**
5.1 Study description . . . 52
5.1.1 Description of the attributes and levels . . . 53
5.2 Business questions . . . 54
5.3 Old estimation procedure . . . 54
5.4 The new methodology . . . 55
5.5 Simulation procedures . . . 56
5.5.1 Improvement over old estimation procedure . . . 57

6 Bundles **59**
6.1 the Bundle study . . . 60
6.2 Differences and similarities with optional features . . . 60
6.2.1 Interaction effect . . . 61
6.2.2 Goals of the model . . . 61
6.2.3 Limitations of the model . . . 61
6.3 Utility from a business perspective . . . 62

7 Results of the Bundles study **63**
7.0.1 Characteristics of the study . . . 63
7.0.2 Choice task design . . . 67
7.0.3 Attributes . . . 68
7.0.4 The choice tasks . . . 70
7.0.5 Results . . . 74
7.0.6 Comment on the coding . . . 75

8 CBC HB estimation algorithm in R **83**
8.1 Running the program . . . 83
8.1.1 Design matrix and answers . . . 84
8.1.2 Prior parameters . . . 85
8.1.3 Running the algorithm . . . 85
8.2 Output files . . . 86

Abstract

The topic of this Master Thesis is quantitative methodologies for optional features and bundles. The frame is the one of Quantitative Marketing Research, a field whose goal is to give market intelligence in forms of, among others, market shares, population clustering and scenario simulations. The particular problem we have worked on is the one of optional features and bundles i.e. services that can be selected for an extra price when purchasing a product.
The technique we have used in our analysis is a discrete choice model, Choice-based Conjoint.

The content of this thesis is based on an internship at the international market research company SKIM. The internship was jointly supervised by Senior Methodologist Kees van der Wagt (SKIM) and Prof. Dr. Richard Gill (Mathematisch Instituut Leiden).

The two most important results of the thesis are new methodologies to study products with optional features and bundles. These methodologies produce utilities that match the respondent's observed choices. Only knowing the estimated utilities, we are able to answer the questionnaire producing answers similar to the observed ones.
The methodologies enjoy all typical properties of conjoint methodologies and can be used to calculate market shares, simulate scenarios etc.
Their most interesting feature is that it is possible to tell if offering an option makes a product too complicated. They can also tell if their simple presence makes the product more appealing (halo effect).

As far as we know, this is the first study in this promising field.
The methodologies we propose are tested on two different datasets arising from studies conducted by SKIM. They have been developed with tests on simulated datasets.
The software of choice for the estimation procedure was Sawtooth's implementation of CBC HB. For reproducibility of experiments we also wrote a package in the open source language R reproducing the same algorithm. This package and Matlab codes used in simulations are found in the Appendix.

Chapter 1

Quantitative techniques for marketing

Market research is the discipline of analyzing and exploring markets. The goal is to acquire valuable information that can be used in taking strategic marketing decisions.
The scope of market research is extremely wide and, depending on the kind of decision to be taken extremely different techniques can be used.

For example, a firm in the automotive industry may be interested in forecasting the state of the market in the following years. They may be interested in knowing how consumers respond to their advertisement. Are the cars they produce a status symbol? What kind of feeling do they elicit in customers? What kind of people their customers are, in terms of age, income, sex? And who are them, in terms of aspirations, values, and dreams? They could be interested in a precise financial forecast of the sale of a new model. How people would respond to a completely different kind of cars being launched on the market? What kind of options should they offer with their cars? Is their offering of cars balanced? What is the optimal price for their line of products? All these questions fall within the scope of market research. They are extremely different, and need completely different methodologies to be answered.

At the most general levels, market research can be divided in two kinds: qualitative and quantitative. Qualitative market research is focused on understanding customers by considering them singularly. The goal is to understand what drives people in their choices or what their perception is of a certain brand or product. Often qualitative research involves panels and in depth discussion about perceived characteristics of a product with a restricted number of study subjects. The goal of qualitative techniques is to give a deep market understanding. In this sense, qualitative methodologies are useful to shape a strategy but are not per se a tool to take decisions.

Quantitative techniques usually provide market understanding based on sound data. The goal is to provide financial forecasts, market shares calcu-

lations, scenario analysis, clusterings of populations. These methodologies are especially good for taking strategic decisions. Usually quantitative techniques rely strongly on statistical techniques to give robust results.

1.1 Conjoint methodologies

Conjoint methodologies are a particular kind of methodologies for quantitative market research. They are based on direct data collection: data are collected especially for each study and no historical data is used.
Further more, they are based on experiments: research participants have to complete carefully engineered exercises that will show their buying behavior.

The roots of conjoint methodologies lie in experimental psychology and psychometrics. This techniques have been used in marketing since 1980s.

In the conjoint experiments respondents have to consider a finite set of products and state their preference(s) in form of choices, numerical ratings or ordering best-to-worst.

The products shown are usually described by a list of their features and sometimes by a picture. Those can be features already present on the market or new features that will be introduced in the future. The ability to study reactions to new features is a strong point for conjoint methodology. It makes possible to study scenarios in which completely new products are introduced. No methodology based on historical data can do such a thing.

The name conjoint is derived from the fact that data is obtained by showing respondents a situation in which they have to evaluate a product in its integrity. Therefore, their preferences for single elements of that product are considered jointly.
In other methodologies respondent may be asked to consider attributes one by one and evaluate them. For example they could be asked to indicate how important for them is to have a GPS included in the price, or what is the maximum price they would consider paying for a given model of car. In conjoint methodologies, respondents only state preferences about full products. Therefore, their preferences (for attributes) are considered jointly. From these joint preferences it is possible to work out the preferences for single attributes and the trade-off between different attributes.

Conjoint methods are at the moment the standard of the market and they have huge advantages over other methods present on the market. First of all, in market research data availability is one of the most critical issues.

Collecting the right data is always difficult and decisions based on a biased or senseless dataset can be disastrous. Many techniques make predictions by looking at historical data and past trends and then extrapolating the results into the future. These techniques can capture trends but are of no use when a completely new product enters the market. Also, it is very hard or sometimes impossible to access sales time series for competitor's single products.

Other techniques are based on data collected at market points (supermarkets, shops etc.). It is very hard to tell if this data is really representative of the whole sample, and it is almost impossible to tell what will happen in case a new product is launched on the market.

The strength of conjoint analysis with respect to such techniques is that it is based on primary data collection. Data is gathered for the specific need of the study. For this reason, the researcher can control the way the random sample is generated and can ask questions of interest.

Another great advantage of conjoint techniques is realism. Conjoint choice tasks usually involve the choice between a number of products that are shown in their integrity. This is very similar to an actual choice purchase and therefore it is a not very demanding task. Also, realism in the choice task gives realistic answers. Ratings given when considering a single feature can be extremely misleading. It is a well known fact that such self-explicated preferences can be very unrealistic. Most people are not able to work out the importance they give to a single attribute. Generally, people tend to state many features are must-have: so important that they will not consider products without them. In reality, most of them are willing to make a trade-off.

Conjoint exercises are mostly of three kinds. Traditionally it was administered as a ranking exercise, when the respondent has to rate some products from most interesting to least interesting. It could also be a rating exercise (where the respondent awards each trade-off scenario a score indicating appeal).

In more recent years it has become common practice to present the trade-offs as a choice exercise. The respondent simply chooses the most preferred alternative from a selection of competing alternatives. This is the most realistic type of exercise, since it mimics actual behavior in the market. In case of a choice exercise, we speak of Choice-Based Conjoint or CBC.

A special kind of choice exercises are constant sum allocation exercises. Respondents are asked to allocate a fixed number of purchases among a set of products. This is meant to represent a series of purchases. The respondent is free to select (i.e. buy) a single product as many times as he/she wants. This kind of exercise is appropriate for products for which consumers show a variety searching behavior. It is also particularly common in pharmaceutical market research, where physicians are given a patient description and have to specify how often are they going to prescribe each of the alternatives. In this case each alternative is the description a real or hypothetical drug/therapy.
For the point of view of estimation, each allocation is considered as independent from the other. Therefore, an allocation exercise with a total sum of 5 is considered as 5 independent CBC questions. This mean that the same estimation procedure for CBC can be used.

Conjoint estimation is traditionally carried out with some form of multiple regression model, but more recently the use of hierarchical Bayesian analysis has become widespread, enabling the study of data at a respondent's level.

Chapter 2

Choice-based Conjoint

2.1 What is CBC conjoint

Choice-based conjoint (CBC) is a particular kind of conjoint methodology. In CBC experiments respondents are shown a certain number of products and they are asked to choose the one they would buy.

Figure 2.1: Example of a CBC choice task

The main advantage of CBC experiments is realism: the task respondents are asked to perform is the same as the actual decision they take when making a purchase.

The goal of CBC conjoint studies is to estimate preferences respondents have for the various features. These preferences are described numerically, forming a set of utilities. As common practice in microeconomics, utility is a numerical value representing the satisfaction that a person receives from a certain service or product. The higher the utility, the better. It is a common assumption that people tend to maximize their utility when making a choice.

From a mathematical point of view, CBC is part of the family of discrete choice models. Those are econometrics models describing a choice among a discrete, finite set in terms of utility. We'll describe those models from a formal point of view, specifying their structure in mathematical terms and illustrating the estimation procedure, in section 2.2.

First we will explain the phases of a conjoint study and the assumption and limitations of CBC models.

2.1.1 Phases of a CBC conjoint study

These are five phases in a typical CBC conjoint study:

1. Problem definition and questionnaire generation
2. Screening
3. Data collection
4. Estimation
5. Follow-up

1. Problem definition and questionnaire generation

The first phase of a conjoint study is to define the characteristics of the market that is to be studied and the business questions that one wants to answer.

The first and most important phase is to decide what attributes should be included in the product description.
It is important to choose which attributes are considered in the study: if the description of a product has too many attributes, respondents will not consider all of them but will place importance to only a few. This is known as a simplification strategy. The use of such strategies by the respondent can greatly impair the estimation procedure: conjoint methodologies are based on the fact that all attributes have a weight in making decisions.

The number of question is also important: choosing too many questions per respondent will make the questionnaire more tiring to answer and respondents will start to give senseless answers. Experienced market researcher advise to limit the number of CBC chioce tasks under 14.

After the problem is defined, a different questionnaire is generated for each respondent.
The goal of having different questionnaires is to show all possible levels combinations.

2. Screening

Often a screening is performed on the respondents. Different screening procedures can be applied: one procedure can be the screening performed ex ante,

before the respondent will answer any of the questions of the survey. Another procedure is the screening performed during the questionnaire answering. Usually CBC studies have many demographic questions before the actual CBC exercise starts.

In general the first type of screening happens when the study is set up and the questionnaire is sent to respondents.
This type of screening is easier to control by the person who set up the experiment, because sending the questionnaire to a certain sample with specific characteristics depends only on the decision of the study maker. The goal of ex-ante screening is to obtain a representative sample or a population with certain characteristics.
For example: in evaluating a new tofu based product, we may want half of the population to be vegetarians and half non-vegetarians, because we know that for this class of products the market is segmented in such a way.

The second type of screening in general happens during the first part of the questionnaire, and is often based on demographic information. This type of screening can be less controlled by the study maker, since it depends totally on the way respondents answer to the questions and therefore it is potentially more subject to bias. For example, for a product for teenagers, we may want to screen out respondents that declare their age to be over a certain threshold.

However, it must be take into account that, if the group of people being studied has any form of control over whether to participate in the study, the so called self-selection bias may arise. Indeed participants' decision to participate might be correlated with traits that affect the study making the participants a non representative sample. Self-selection bias is hard to track down and can be the cause of very biased results.

3. Data collection

Respondents answer the questionnaire. The questionnaire can be an actual paper module to be compiled. Today, more and more questionnaires are completed on the internet. Internet questionnaires are cheaper, more time-effective and they are getting increasingly popular.
However, internet surveys are known for generating less precise answers and the number of fraudulent respondents (respondents answering casually) is much higher than with paper and pencil surveys.
Usually data such as the time spent to compile the questionnaire is used to screen out too fast respondents.

4. Estimation

Based on the collected data, utilities are estimated.

Using different algorithms, it is possible to estimate utilities for the whole population considered as one, for homogenous groups of respondents and also for each respondent.
We will explain in detail the estimation procedure in the next sections.

5. Follow up

The estimated utilities are used to develop market insight.
Given the utilities, it is possible to calculate market shares for different products. It is possible to set the current market situation as the base scenario and see how shares change when a change is introduced in the market.
For some populations clearly defined segments may be present, and it is possible to track them down, dividing the population in homogenous groups with similar tastes.
It is possible to study interaction between certain attributes. We can calculate price sensitivity curves at an aggregate, group or respondent level. We can see how the value of a brand is perceived among respondents, and what features have more weight in the choice decision.

2.1.2 Different estimation procedures

As we mentioned earlier, there are different ways to analyze the data collected in a conjoint study. The most important methods are Aggregate Logit, Latent Class and CBC HB. They respectively provide utilities for the whole respondents population, for homogeneous groups in the population and for each component of the population.

Aggregate logit

In this model the whole population of respondents is considered in its integrity. The result is a single set of utilities for the whole population. Intuitively, the resulting utilities describe an average of the population preferences.

Aggregate estimation assumes that the respondent utility is equal to the average utility, which is a quite restrictive assumption and does not allow for idiosyncratic, individual effects in the sample, meaning that heterogeneity in the sample is simply not considered. This was the first model ever implemented to analyze conjoint data. We will explain in detail this model in section 2.2.1.

Latent Class

Cluster analysis is, historically speaking, the evolution of aggregate estimation. It was developed to allow for some form of respondents heterogeneity.

Clustering algorithms find groups of individuals with similar tastes among the whole sample. The preferences of the individuals are estimated in a "semi individual" way by assuming that the respondent utility is equal to the cluster utility, allowing for heterogeneities across segments of respondents but not within the cluster. To allow for heterogeneity between single respondents, HB models were created.

Clustering models are still very important in their own respect. From a commercial point of view it is very important to divide the market in segments that have similar tastes.

Latent Class estimation detects subgroups of respondents with similar preferences and estimates utilities for each segment. Each respondent is given a probability to being part of a certain group.
It is possible to specify how many groups are to be considered. There are criteria (notably the Akaike criterion) to decide the optimal number of groups to consider.

CBC HB

HB methods are the newest and currently most used estimation methods in quantitative marketing research.
The name CBC HB means "Choice Based Conjoint - Hierarchical Bayes". The mathematical specification of these model is a Bayesian hierarchical model in which, broadly speaking, a different vector of utility is define for each respondent. The distribution of these utilities in the whole population has some specified form, usually normal.

CBC HB allows for heterogeneity at a respondent's level by specifying different utilities for each respondent. This leads to a greater improvement in simulation techniques: simulation conducted using aggregate or clusterized models often lead to biased results.
We will devote most of chapter 3 to explain the details of this model.

2.1.3 A Study example

To make things clearer, we show what the result of a CBC study would look like.
Suppose we are interested in the market of smartphones.
We think the most important features are of course price, then brand, screen size, internal memory, operative system and if there is a keyboard or a touch screen. These are by no mean all attributes, but we must limit the number of attribute to study to have meaningful answers from the respondents.

Considering some phones on the market, we decide to these attributes can take the following values:

- Price: 220 euro, 230 euro, ...490 euro, 500 euro
- Brand: Nokia, Samsung, Blackberry, LG
- Screen size: 2.8", 3", 3.2", 3.5"
- Internal memory: 2 Gb, 4 Gb, 8 Gb, 16 Gb
- Full keyboard: present/not present
- Touch screen: present/not present

For example a Nokia E63, a model really present on the market, is defined by the following vector of attribute levels: (240 euro, Nokia, 3", 4 GB, present, not present).

A choice task with 4 alternatives is a list of 4 configuration vectors. These don't need to represent any phone really present on the market and ca be completely casual. A typical choice task would look something like

Choose one of the following

	Product 1	Product 2	Product 3	Product 4
Price	300	250	230	370
Brand	Nokia	Blackberry	Samsung	LG
Screen size	2.8"	3.2"	3.2"	3"
Internal Memory	2Gb	4Gb	8Gb	8Gb
Full Keyboard	present	not present	present	not present
Touch screen	not present	present	present	present

After collecting the answers to the choice tasks, we can perform the estimation.
If using a the aggregate logit model, we will obtain a single vectory of utilities. We have an utility for each level. For each attribute will we obtain the utilities:

	Brand:Nokia	Brand:Samsung	Brand:LG	Brand:Blackberry
Utility	2.18	0.4	3.24	-0.76

	Screen size: 2.8	Screen size: 3	Screen size: 3.2	Screen size: 3.5
Utility	4.30	-0.74	-1.64	2.96

and so on.

In case we used CBC-HB, we would have different utilities for each respondent:

	Brand:Nokia	Brand:Samsung	Brand:LG	Brand:Blackberry
Resp. 1	2.18	0.4	3.24	-0.76
Resp. 2	4.23	-0.82	1.96	2.28
...	...	...	...	...
Resp. 300	2.193	1.12	2.93	-1.20

Suppose we are interested in the market shares in a given scenario. For scenario we intend the situation of available products on the market.

First we define the products available on the market. Just for the example's sake, we consider a market with only 4 products.

Market situation

	Product 1	Product 2	Product 3	Product 4
Price	300	250	230	370
Brand	Nokia	Blackberry	Samsung	LG
Screen size	2.8”	3.2”	3.2”	3”
Internal Memory	2Gb	4Gb	8Gb	8Gb
Full Keyboard	present	not present	present	not present
Touch screen	not present	present	present	present

To calculate the market share of each phone, we first calculate each phone's utility. The utility of the phone is just the sum of its features utilities.

Let's first consider the case of Aggregate Logit. This is what we may find:

	Product 1	Product 2	Product 3	Product 4
Utility	10,23	11,67	4.48	9,24
Exp(Utility)	27722	117008	73130	10301
Market share	12,5%	51%	32%	4,5%

To calculate market shares, we use a method known as share of preference. In this method, we first exponentiate each product utility. The market share of one product is that product exponentiated utility divided by the sum of all products exponentiated utilities.
This way, the market share of each product is proportional to its exponentiated utility.

If we used the CBC HB algorithm, we can calculate product utilities for each respondent.

Utility per respondent	Product 1	Product 2	Product 3	Product 4
Resp 1	10,23	11,67	4.48	9,24
...	...	...	...	...
Resp 300	16,76	12,67	14.78	12,5

In this case to calculate market shares we can use again share of preference, calculating a vector of market shares for each respondent. The final market share of a product is the average of the market shares calculated for each respondent. However, it's also possible to calculate shares in a different way. We imagine each respondent is making a single choice, and we assume he/she would chose the product with the highest utility.
So for example Respondent 1 would choose product 2 and respondent 300 would choose product 1.
At the end the market share of a product is the number of times it was chosen divided by the number of respondents. This method of calculating market shares is called First Choice.

We are now going to explain the reasons behind the use of these two methods.

2.1.4 Scenario simulation methods

The ability to perform scenario simulations is the most interesting feature of conjoint studies.
Using the estimated utilities,we can calculate market shares of the products present now on the market. We can see what happens when an existing product changes its design or price, or when a new product enters the market.
Further more, it is possible to see what would happen if a completely new class of products enters the market.

The most interesting simulations are calculated with CBC-HB models, where we can simulate the choice of every single respondent.

There are mainly three methods to simulate choice, namely first choice, share of preference and randomized first choice.
First choice assumes that each respondent chooses one product (the one maximizing utility), while share of preference (also called logit simulation) indicates for each respondent a share of purchase for each product, proportional to their utilities.
Randomized first choice is somehow an alternative in between the two previous methods: each respondent chooses a single product with a probability proportional to their utilities.

In first choice of preference the product with the highest utility is chosen by the respondent, and then the share of products across respondents is calculated by dividing the number of respondents choosing that product by the total respondents. This method relies on the assumption that the respondent doesn't have a variety seeking behavior and spends enough time considering the purchase, being able to identity the product with the highest utility.

This method makes sense for purchase decisions that respondents evaluate carefully and that involves huge amounts of money, such as the purchase of a car or a house.

Share of preference method assumes a probability distribution across products for each respondent. For each respondent, the exponential of each product's utility, divided by a normalizing constant, is the share of preference of that product for that respondent.
This method makes sense for products where several purchases are made in a certain period. For example, when considering products like jam consumer do not always buy the one they like most and they are likely to buy different tastes in different purchases.
Since it is based on a distribution of preferences, share of preference provides a flatter distribution of shares with respect to first choice.
Share of preference is to be preferred for those product whose purchase is not very carefully considered or for which there may be a variety seeking behavior. Examples are CPG products like biscuits, chips, soft drinks etc.

2.1.5 Limitations of CBC

The CBC model is such a powerful and easy tool that one can be seduced into using it to analyze all marketing problems. Indeed, despite (or we could say, because of) its great scope of analysis, the CBC model has to be used with care.

The first limitation is not a limitation of the model itself, but rather of the respondents. Cognitive psychology experiments have confirmed what marketers knew from long time: people can answer a very limited number of choice tasks before loosing interest and motivation and answering just randomly.
As a rule of thumb, no more than 14 questions should be asked to each respondent in a conjoint study. A common way to discriminate well thought answers from random clicks is to analyze the amount of time a respondent spends analyzing each question.

The amount of time spent for each choice task diminishes invariably after each question. This can be explained both by a progressive lack of motivation and by the fact that respondents get experience in performing choice tasks.
There is evidence that well motivated respondents take less time to answer choice tasks as they get more accustomed to the characteristics of the products shown.

Another factor to consider is the number of features: each product should be made of a number of attributes limited enough that the respondent can consider all of them at once. If this is not the case, respondents will apply simplification strategies: they will base their choice only on some of the attribute shown, considering the other unimportant. This gives unrealistic results.
It is advisable not to use more than 7 features.

In the specification of the model, attributes evaluation of one product should be independent of each other. For example, in a car study the utility for a single respondent for the color blue or red should be independent from the brand of the car.
However, what happens in reality is that for most people the color Red is more appealing when offered with a Ferrari than with a Porsche. That is, there is an interaction between two attributes.

For most products an assumption of independent features is not always realistic and sometimes is just wrong. It is possible to estimate the interaction effect between two attributes.

If an interaction effect is known to be present between two attributes, a common practice is to add the two attributes in one.
In the setting of the previous example, we would start with the attributes
Brand: Ferrari, Porsche, Jaguar (3 levels) Color: Red, Dark green, Silver (3 levels)
In this case, the results may show people having a strong preference for silver

or red or dark green. So the model would tell us that a Silver Ferrari would be better than a Red Ferrari, something we know is not true.

The solution is to consider a new attribute made of all the combinations of the previous ones: Brand+color: Red Ferrari, Dark green Ferrari, Silver Ferrari, Red Porsche, Dark green Porsche, Silver Porsche, Red Jaguar, Dark green Jaguar, Silver Jaguar (9 levels)
In this way we would be able to see correctly that the utility for a Red Ferrari would be much better than the one for a Dark Green Ferrari and so on.

In less extreme cases, it is possible to ignore the interaction. Market shares are quite consistent even in this case.

It is a well know fact by marketers that people can show much different price sensitivity in real market than in choice experiments.
For example, when buying a bag of chips most people don't really spend too much time considering the price, given it is in an "acceptable" range and just pick one bag that has a taste they like.
When performing a choice task in a conjoint experiment about chips, the same person may be taken to consider its purchase in a much more precise way. Therefore he/she may show much higher price sensitivity.
There is no real way to see if the results show too much price sensitivity other than having some experience with one market. The best way to solve this is to re-scale the price utilities by some constant and then doing simulations as always.

When studying a market scenario, one may be interested in comparing the predicted shares of the current scenario to the ones measured on the market. In doing so, one could be quite disappointed as the estimated market shares may be quite far from the actual ones.

This is not a reason for concern: the results from a conjoint study may be perfectly reasonable and very worthy even when they are not able to predict the current market share.
This is because the conjoint market shares are calculated in a much idealized condition: all customers have perfect information about the products, all the products are accessible to all consumers and the customers are driven in their choice only by the features of the product rather than promotions, advertisement campaigns and so on.
So for example, from a CBC study on cigarettes we expect the resulting market shares to be extremely precise since there are few products on the market and they are pretty much available everywhere.

For other products like cars it may be more complicated: respondents may show great interest in a particular Hyundai model, giving a simulated market share much higher than the real one. This may be due to the lack of Hyundai sellers in the country, or to some aggressive price promotion by competitors, or to public founding for certain type of cars.
This doesn't mean that the conjoint results are not useful.

When a firm has to take a strategic decision it must compare the market

share in the current (simulated) scenario and the one in a new scenario. The only thing that matters is the difference between the two.
Since conjoint studies deliver ideal shares, they filter all external influences on market share.

In this sense, ideal market shares are even better to understand modifications in the market scenario. This is even better when taking a strategic decision.
If the aim is to get a precise financial forecast, it is possible to correct ideal market share to take into account external factors.

When considering the scenario results of a conjoint study (for example a market share for a certain product) it must also be noted that those are peak results.
For example, even for a perfect representation of people's tastes and in a market scenario that will not change in the immediate future, the market share of a new product will not be achieved right away after its launch.

It takes time for people to switch from one product to the new one and typically you have to educate people about the features of the new product and some people just take a lot of time to decide to switch.
To estimate the amount of time needed to reach this peak share a general knowledge of the market is needed. In some markets it takes a great deal of time to convince people to switch from the product they are using, while in others it is way easier.

Variety searching is a well documented behavior, both in marketing practice and econometric theory. When making a choice, customers do not always choose their ideal but like to try other products. This is especially true for consumer packaged goods (CPG), inexpensive products that are purchased regularly like food, drinks etc.

The opposite of variety searching is habit forming: the tendency of customers to stick to the product they are already buying even when a new product, closer to their ideal, enters the market.

Furthermore, certain products may have some sort of barrier to change. A user of a certain product may encounter trouble when wanting to switch to another one because of the existence of a contract, or because of costs to be taken when switching.
Think for example of an office that wanted to switch from Microsoft Windows to an Apple OS: most of their software licenses would be useless and they would have to change most of the hardware.

CBC market shares don't account for all this issues: if a new product is featured in a scenario, its market share will not consider barriers to change and previous habits of customers.
To conclude, CBC shares are much idealized: they are calculated as if customers

had perfect knowledge of all the alternatives and had no history, as if they were in the market for the first time and not locked to any brand or product.

2.2 Discrete choice models

Discrete choice models are models used in econometrics to describe choices by rational actors in a finite set.
These models describe the choice as depending from some observable characteristics of the elements in the set and some parameters, unknown to the researcher, called utilities.
The model used for conjoint studies is also called logit model and is special cases of discrete choice models.

2.2.1 Derivation of choice probabilities

Utility represents the benefit gained when selecting an element of the set.
Discrete choice models are usually derived under an assumption of utility-maximizing behavior by the decision maker.

A decision maker, labeled n, faces a choice among J alternatives. We assume the respondent would obtain a certain level of utility from each alternative. The utility that decision maker n obtains from alternative j is $U_{nj}, j = 1 \dots J$.

This utility is known to the decision maker but not by the researcher. The decision maker chooses the alternative that provides the greatest utility. The behavioral model is therefore: choose alternative i if and only if $U_{ni} > U_{nj} \forall j \neq i$.

From the researcher's point of view, it is not possible to observe the decision maker's utility. The only thing that can be observed is some attributes of the alternatives, which we will call x_j. It is also possible to define some attribute of the decision maker, called β_n and to specify a function that related these attributes to the decision maker utility. This function is denoted $V_{\mathrm{nj}} = V(x_{\mathrm{nj}}, \beta_n) \forall j$ and it is usually called representative utility.

Utility is decomposed as $U_{\mathrm{nj}} = V_{\mathrm{nj}} + \varepsilon_{\mathrm{nj}}$, where $\varepsilon_{\mathrm{nj}}$ captures the factors that affect utility but are not included in V_{nj} . This decomposition is fully general, since $\varepsilon_{\mathrm{nj}}$ is defined as simply the difference between true utility U_{nj} and the part of utility that the researcher captures in V_{nj}.

Given this definition, the characteristics of $\varepsilon_{\mathrm{nj}}$, such as its distribution, depend critically on the researcher's specification of V_{nj}.

Usually the researcher specifies the analytical form of V_{nj} and $\varepsilon_{\mathrm{nj}}$ according to a model describing his/her assumptions about the choice. The x_{nj}'s are usually considered known (as they describe the choice alternatives) and the interest of the researcher is to find values of the parameters β_n that in some sense better describe the observed choices.

The researcher does not know $\varepsilon_{\mathrm{nj}} \forall j$ and therefore treats these terms as random. They are usually called error terms. The joint density of the random

vector $\varepsilon_n = (\varepsilon_{n1}, \ldots \varepsilon_{nj})$ is denoted $f(\varepsilon_n)$.Knowing the density, the researcher can make probabilistic statements about the decision maker's choice. The probability that decision maker n chooses alternative i is

$$\begin{aligned} P_{\text{ni}} &= P(U_{\text{ni}} > U_{\text{nj}}, \forall j \neq i) \\ &= P(V_{\text{ni}} + \varepsilon_{\text{ni}} > V_{\text{nj}} + \varepsilon_{\text{nj}}, \forall j \neq i) \\ &= P(\varepsilon_{\text{nj}} - \varepsilon_{\text{ni}} < V_{\text{ni}} - V_{\text{nj}}, \forall j \neq i) \end{aligned} \tag{2.1}$$

This probability is a cumulative distribution, namely, the probability that each random term $\varepsilon_{\text{nj}} - \varepsilon_{\text{ni}}$ is below the observed quantity $V_{\text{ni}} - V_{\text{nj}}$.

Using the density $f(\varepsilon_n)$ the cumulative probability can be rewritten as

$$P_{\text{ni}} = P(\varepsilon_{\text{nj}} - \varepsilon_{\text{ni}} < V_{\text{ni}} - V_{\text{nj}}, \forall j \neq i)$$

$$= \int_{\varepsilon} I(\varepsilon_{\text{nj}} - \varepsilon_{\text{ni}} < V_{\text{ni}} - V_{\text{nj}}, \forall j \neq i) f(\varepsilon_n) d\varepsilon_n$$

where I(.) is the indicator function, equaling 1 when the expression in parenthesis is true and 0 otherwise. This is a multidimensional integral over the density of the unobserved portion of utility, $f(\varepsilon_n)$. Different discrete choice models are obtained from different specifications of this density, that is, from different assumptions about the distribution of the unobserved portion of utility. The integral takes a closed form only for certain specifications of $f(.)$. Logit models have closed form expressions for this integral.

2.2.2 Utilities and additive constants

If a constant is added to the utility of all the alternatives, the alternative with the highest utility doesn't change. Since the respondent always chooses the alternative with the highest utility, the choice is the same with $U_{\text{nj}} \forall j$ as with $U_{\text{nj}} + k \forall j$ for any constant k. Therefore from the respondent's point of view, the absolute value of the utility is meaningless and the only thing that counts is the difference with the other utilities.

Things don't change from the researcher's perspective. The choice probability is $P_{\text{ni}} = P(U_{\text{ni}} > U_{\text{nj}}, \forall j \neq i) = P(U_{\text{ni}} - U_{\text{nj}} > 0, \forall j \neq i)$, which depends only on the difference in utility, not its absolute level.

When utility is decomposed into the observed and unobserved parts, equation 2.1 expresses the choice probability as $P_{\text{ni}} = P(\varepsilon_{\text{nj}} - \varepsilon_{\text{ni}} < V_{\text{ni}} - V_{\text{nj}}, \forall j \neq i)$ which also depends only on differences between utilities.

Therefore, since utilities are defined up to an additive constant, the absolute value of utility can not be estimated, as there are different sets of utilities leading to the same choices. This must be taken into consideration when comparing two sets of utilities.

2.2.3 Utility scale

We have seen that adding a constant to all utilities doesn't change respondent's behavior as the alternative with the highest utility doesn't change. The same happens when multiplying all utilities for a given positive constant.

The model $U^0_{\text{nj}} = V_{\text{nj}} + \varepsilon_{\text{nj}}$ is equivalent to $U^1_{\text{nj}} = \lambda V_{\text{nj}} + \lambda\varepsilon_{\text{nj}}$ for any $\lambda > 0$: the alternative with the highest utility is the same no matter how utility is scaled. To take account of this fact, we have to normalize the scale of utility.

We can normalize the scale of utility by normalizing the variance of the error term. When utility is multiplied by λ, the variance of each ε_{nj} is multiplied by λ^2: $\text{var}(\lambda\varepsilon_{\text{nj}}) = \lambda^2 \text{var}(\varepsilon_{\text{nj}})$.

When error terms are assumed to be i.i.d. (as it is for most models) it is easy to normalize the error variance of all terms setting it to some value usually chosen for convenience.

The error variances in a standard logit model are usually normalized to $\frac{\pi^2}{6}$. In this case, the model becomes $U_{\text{nj}} = x'_{\text{nj}}(\beta/\sigma) + \varepsilon_{\text{nj}}/\sigma$ with $\text{var}(\varepsilon_{\text{nj}}) = \frac{\pi^2}{6}\sigma$.

2.3 Logit Model

2.3.1 Choice probabilities

Let's consider again the general discrete choice model in which decision maker n chooses among J alternatives.

The utility of a given alternative j is decomposed into a part labeled V_{nj} that is known by the researcher up to some parameters and an unknown part ε_{nj} (the error term) that is treated by the researcher as random: $U_{\text{nj}} = V_{\text{nj}} + \varepsilon_{\text{nj}} \forall j$. The logit model is obtained by assuming that each ε_{nj} is independently identically distributed Gumbel with location parameter $\mu = 0$.

The density for each unobserved component of utility is

$$f(\varepsilon_{\text{nj}}) = \frac{e^{-\varepsilon_{\text{nj}}}}{\sigma} e^{\frac{-\varepsilon_{\text{nj}}-\mu}{\sigma}}$$

and the cumulative distribution is

$$F(\varepsilon_{\text{nj}}) = \exp(-\exp(-\frac{\varepsilon_{\text{nj}} - \mu}{\sigma})) \tag{2.2}$$

The variance of this distribution is $\frac{\pi^2}{6}\sigma$.

To normalize the scale of utility the variance of the ε_{nj} terms is set to the standard values $\frac{\pi^2}{6}$ by dividing U_{nj} by σ:

The most important feature of the Gumbel distribution is that the difference between two i.i.d. Gumbel variables has a logistic distribution.

Theorem 2.3.1 *If ε_{nj} and ε_{ni} are i.i.d. Gumbel, then $\varepsilon^* = \varepsilon_{\text{nj}} - \varepsilon_{\text{ni}}$ follows the logistic distribution. This distribution has CDF:*

$$F(\varepsilon^*_{\text{nji}}) = \frac{e^{\varepsilon^*_{\text{nji}}}}{1 + e^{\varepsilon^*_{\text{nji}}}}$$

The assumption that errors are independent of each other is very important and could be seen as restrictive.
Actually, it should be seen as the outcome of a well-specified model. The error term ε_{nj} is just the unobserved portion of utility for one alternative. This is defined as the difference between the utility that the decision maker actually obtains, U_{nj}, and the representation of utility that the researcher has developed using observed variables, V_{nj} .

Under independence, the unobserved portion for one alternative provides no information to the researcher about the unobserved portion of another alternative. Stated equivalently, the researcher has specified the form of the representative utility with such a degree of precision that the remaining, unobserved portion of utility is essentially noise: all the needed information relevant in the decision process is captured in the analytical form of V_{nj} .

In a deep sense, the ultimate goal of the researcher is to represent utility so well that the only remaining aspects constitute simply white noise; that is, the goal is to specify utility well enough that a logit model is appropriate.

We now derive the logit choice probabilities, following McFadden (1974). The probability that the decision maker n chooses alternative i is

$$\begin{aligned} P_{\text{ni}} &= P(V_{\text{ni}} + \varepsilon_{\text{ni}} > V_{\text{nj}} + \varepsilon_{\text{nj}}, \forall j \neq i) \\ &= P(\varepsilon_{\text{nj}} < \varepsilon_{\text{ni}} + V_{\text{ni}} - V_{\text{nj}}, \forall j \neq i) \end{aligned} \tag{2.3}$$

For each j, the cumulative distribution of ε_{nj} evaluated at $\varepsilon_{\text{ni}} + V_{\text{ni}} - V_{\text{nj}}$ is:

$$F_{\varepsilon_{\text{nj}}}(\varepsilon_{\text{ni}} + V_{\text{ni}} - V_{\text{nj}}) = \exp(-\exp(-(\varepsilon_{\text{ni}} + V_{\text{ni}} - V_{\text{nj}})))$$

Let's call $P_{\text{ni}}(\varepsilon_{\text{ni}})$ the value of the probability P_{ni} given the value of ε_{ni}. Since ε's are independent, this probability over all $j \neq i$ is the product of the individual cumulative distribution

$$P_{\text{ni}}(\varepsilon_{\text{ni}}) = \prod_{j \neq i} \exp(-\exp(-(\varepsilon_{\text{ni}} + V_{\text{ni}} - V_{\text{nj}})))$$

The choice probability P_{ni} is the integral of $P_{\text{ni}}|\varepsilon_{\text{ni}}$ over all values of ε_{ni} weighted by its density 2.2.

$$P_{\text{ni}} = \int (\prod_{j \neq i} \exp(-\exp(-(\varepsilon_{\text{ni}} + V_{\text{ni}} - V_{\text{nj}})))e^{-\varepsilon_{\text{ni}}} e^{-e^{-\varepsilon_{\text{ni}}}} d\varepsilon_{\text{ni}}$$

which we rewrite as

$$P_{\mathrm{ni}} = \int_{s=-\infty}^{+\infty} (\prod_{j \neq i} \exp(-\exp(-(s + V_{\mathrm{ni}} - V_{\mathrm{nj}}))) e^{-s} e^{-e^{-s}} \, \mathrm{ds}$$

where $s = \varepsilon_{\mathrm{ni}}$.

We note that $V_{\mathrm{ni}} - V_{\mathrm{ni}} = 0$. Collecting terms in the exponent of e, we have

$$P_{\mathrm{ni}} = \int_{s=-\infty}^{+\infty} (\prod_j e^{-e^{-(s+V_{\mathrm{ni}}-V_{\mathrm{nj}})}}) e^{-s} \, \mathrm{ds}$$

$$= \int_{s=-\infty}^{+\infty} \exp(-\sum_j e^{-e^{-(s+V_{\mathrm{ni}}-V_{\mathrm{nj}})}}) e^{-s} \, \mathrm{ds}$$

$$= \int_{s=-\infty}^{+\infty} \exp(-e^{-s} \sum_j e^{-(V_{\mathrm{ni}}-V_{\mathrm{nj}})}) e^{-s} \, \mathrm{ds}$$

we rewrite it calling $e^{-s} = t$, with $-e^{-s} \, \mathrm{ds} = \mathrm{dt}$.

$$P_{\mathrm{ni}} = \int_{\infty}^{0} \exp(-t \sum_j e^{-(V_{\mathrm{ni}}-V_{\mathrm{nj}})} (-\, \mathrm{dt})$$

$$= \int_{0}^{\infty} \exp(-t \sum_j e^{e(-V_{\mathrm{ni}}-V_{\mathrm{nj}})}) \, \mathrm{dt}$$

$$= \frac{\exp(-t \sum_j e^{-(V_{\mathrm{ni}}-V_{\mathrm{nj}})}}{-\sum_j e^{-(V_{\mathrm{ni}}-V_{\mathrm{nj}})}} |_0^{\infty}$$

$$= \frac{1}{\sum_j e^{-(V_{\mathrm{ni}}-V_{\mathrm{nj}})}} = \frac{e^{V_{\mathrm{ni}}}}{\sum_j e^{V_{\mathrm{nj}}}}$$

From the integral at the beginning we arrived to this easy closed form expression

$$P_{\mathrm{ni}} = \frac{e^{V_{\mathrm{ni}}}}{\sum_j e^{V_{\mathrm{nj}}}}$$

which is the logit choice probability. The fact that choice probabilities are expressed in a closed form is one of the biggest advantages of logit over other discrete choice models, for example probit. Logit's choice probabilities are faster to compute: to calculate choice probabilities in a n alternative probit model we have to approximate the value of a n-uple integral. This is a great advantage when performing simulation-based estimation. Representative utility is usually specified to be linear in parameters: $V_{\mathrm{nj}} = \beta' x_{\mathrm{nj}}$ where x_{nj} is a vector of observed variables describing alternative j. With this specification, the logit probabilities become

$$P_{\mathrm{ni}} = \frac{e^{\beta' x_{\mathrm{nj}}}}{\sum_j e^{\beta' x_{\mathrm{nj}}}}$$

Another positive feature of linear utilities is that the log-likelihood function with these choice probabilities is globally concave in parameters β, which leads to a unique maximum and a faster optimization. This result can be found in MacFadden (1974).

2.3.2 Estimation procedure

Random sample

A sample of N decision makers, randomly selected across the population, is obtained for the purpose of estimation. Each decision maker has to perform n_{quest} choice tasks, resulting in n_{alt} choices.

In a single choice task with n_{alt} alternatives, the probability of person n choosing the alternative that was actually observed as a choice in question k is

$$\prod_{w=1}^{n_{\text{alt}}} (P_{\text{nwk}})^{y_{\text{nwk}}}$$

where $y_{\text{nik}} = 1$ if person n chose alternative i and 0 otherwise.

For convenience we assume each choice task had the same number of alternatives.

We assume that the choices in different questions are independent from each other. For convenience we assume each choice task had the same number of alternatives. Therefore the probability of observing the actual choices is

$$\prod_{k=1}^{n_{\text{quest}}} \prod_{w=1}^{n_{\text{alt}}} (P_{\text{nwk}})^{y_{\text{nwk}}}$$

To make the calculations easier we can write it as

$$\prod_{k=1}^{n_{\text{quest}}} \prod_{w=1}^{n_{\text{alt}}} (P_{\text{nwk}})^{y_{\text{nwk}}} = \prod_{i=1}^{n_{\text{quest}} n_{\text{alt}}} (P_{\text{ni}})^{y_{\text{ni}}}$$

where P_{ni}, $i = 1 : n_{\text{quest}} \cdot n_{\text{alt}}$ is just the collection of choice probabilities of each alternative in each question, and $y_{\text{ni}} = 1$ if person n chose alternative i and 0 otherwise. Assuming that each decision maker's choices are independent of those of the other decision makers, the probability of each person in the sample performing the observed choices is:

$$L(\beta) = \prod_{n=1}^{N} \prod_{i} (P_{\text{ni}})^{y_{\text{ni}}}$$

where β is a vector containing the parameters of the model. The log-likelihood function is then

$$\mathrm{LL}(\beta) = \sum_{n=1}^{N} \sum_{i} y_{\mathrm{ni}} \ln(P_{\mathrm{ni}}) \tag{2.4}$$

and the estimator is the value of β that maximizes this function. McFadden (1974) shows that $\mathrm{LL}(\beta)$ is globally concave for linear parameters utility. In this case the maximum likelihood estimate is the unique solution of the first order condition

$$\frac{\mathrm{dLL}(\beta)}{d\beta} = 0$$

For convenience, let the representative utility be linear in parameters: $V_{\mathrm{nj}} = \beta' x_{\mathrm{nj}}$. This specification is not actually required for the final result, but it is the one we are going to use in the rest of this thesis and makes the calculations more succinct. Using (3.11) and the formula for the logit probabilities, we'll show that the first-order condition 2.4 becomes

$$\Sigma_n \Sigma_i (y_{\mathrm{ni}} - P_{\mathrm{ni}}) x_{\mathrm{ni}} = 0 \tag{2.5}$$

We start considering the value of the log likelihood:

$$\mathrm{LL}(\beta) = \sum_{n} \sum_{i} y_{\mathrm{ni}} \ln P_{\mathrm{ni}}$$

$$= \sum_{n} \sum_{i} y_{\mathrm{ni}} \ln(\frac{e^{\beta'} x_{\mathrm{ni}}}{\Sigma_j e^{\beta' x_{\mathrm{nj}}}})$$

$$= \sum_{n} \sum_{i} y_{\mathrm{ni}} (\beta' x_{\mathrm{ni}}) - \sum_{n} \sum_{i} y_{\mathrm{ni}} \ln(\Sigma e^{\beta' x_{\mathrm{nj}}})$$

The derivative of the log-likelihood function then becomes

$$\frac{\mathrm{dLL}(\beta)}{d\beta} = \frac{\sum_n \sum_i y_{\mathrm{ni}} (\beta' x_{\mathrm{ni}})}{d\beta} - \frac{\sum_n \sum_i y_{\mathrm{ni}} \ln(\Sigma e^{\beta' x_{\mathrm{nj}}})}{d\beta}$$

$$= \sum_{n} \sum_{i} y_{\mathrm{ni}} x_{\mathrm{ni}} - \sum_{n} \sum_{i} y_{\mathrm{ni}} \sum_{j} P_{\mathrm{nj}} x_{\mathrm{nj}}$$

$$= \sum_{n} \sum_{i} y_{\mathrm{ni}} x_{\mathrm{ni}} - \sum_{n} (\sum_{j} P_{\mathrm{nj}} x_{\mathrm{nj}}) \Sigma_i y_{\mathrm{ni}}$$

$$= \sum_{n} \sum_{i} y_{\mathrm{ni}} x_{\mathrm{ni}} - \sum_{n} (\sum_{j} P_{\mathrm{nj}} x_{\mathrm{nj}})$$

$$= \sum_{n} \sum_{i} (y_{\mathrm{ni}} - P_{\mathrm{ni}}) x_{\mathrm{ni}} \tag{2.6}$$

setting this derivative to 0 gives the first-order condition of 2.5

Rearranging and dividing both sides by N

$$\frac{1}{N}\Sigma_n\Sigma_i y_{\mathrm{ni}}x_{\mathrm{ni}} = \frac{1}{N}\Sigma_n\Sigma_i P_{\mathrm{ni}}x_{\mathrm{ni}} \tag{2.7}$$

This expression is readily interpretable. Let $\bar{x}$ denote the average of x over the alternatives chosen by the sampled individuals: $\bar{x} = \frac{1}{N}\Sigma_n\Sigma_i y_{\mathrm{ni}}x_{\mathrm{ni}}$. Let $\hat{x}$ be the average of x over the predicted choices of the sampled decision makers: $\hat{x}$= (1/N)$\Sigma_n\Sigma_i P_{\mathrm{ni}}x_{\mathrm{ni}}$. The observed average of x in the sample is $\bar{x}$, while $\hat{x}$ is the predicted average. By 2.7, these two averages are equal at the maximum likelihood estimates. That is, the maximum likelihood estimates of β are those that make the predicted average of each explanatory variable equal to the observed average in the sample.

In this sense, the estimates induce the model to reproduce the observed averages in the sample.

An alternative-specific constant is the coefficient of a dummy variable that identifies an alternative. A dummy for alternative j is a variable whose value in the representative utility of alternative i is $d_i^j = 1$ for $i = j$ and zero otherwise. By 2.7, the estimated constant is the one that gives

$$\frac{1}{N}\sum_n\sum_i yni\,\mathrm{d}_i^j = \frac{1}{N}\sum_n\sum_i P_{\mathrm{ni}}d_i^j$$

$$S_j = \widehat{S_j}$$

where S_j is the share of people in the sample who chose alternative j , and $\widehat{S_j}$ is the predicted share for alternative j . With alternative-specific constants, the predicted shares for the sample equal the observed shares. The estimated model is therefore correct on average within the sample.

This feature is similar to the function of a constant in a linear regression model, where the constant assures that the average of the predicted value of the dependent variable equals its observed average in the sample. The first-order condition 2.3.2 provides yet another important interpretation. The difference between a person's actual choice, y_{ni} , and the probability of that choice, P_{ni} , is a modeling error, or residual. The lefthand side of 2.3.2 is the sample covariance of the residuals with the explanatory variables.

The maximum likelihood estimates are therefore the values of the $\beta's$ that make this covariance zero, that is, make the residuals uncorrelated with the explanatory variables. This condition for logit estimates is the same as applies in linear regression models. For a regression model $y_n = \beta' x_n + \varepsilon_n$, the ordinary least squares estimates are the values of β that set $\Sigma_n(y_n - \beta' x_n) = 0$. This fact is verified by solving for $\beta : \beta = (\Sigma_n x_n x_n')^{-1}(\Sigma_n x_n y_n)$ which is the formula for the ordinary least square estimator. Since $y_n - \beta' x_n$ is the residual in the regression model, the estimates make the residuals uncorrelated with the explanatory variables.

Under this interpretation, the estimates can be motivated as providing a sample analog to population characteristics.We have assumed that the explana-

tory variables are exogenous, meaning that they are uncorrelated in the population with the model errors. Since the variables and errors are uncorrelated in the population, it makes sense to choose estimates that make the variables and residuals uncorrelated in the sample. The estimates do exactly that: they provide a model that reproduces in the sample the zero covariances that occur in the population.

2.3.3 Choice among a subset of alternatives

In some case, the number of alternatives facing the decision maker is so large that estimating the model parameters is computationally very expensive or even impossible.

With a logit model, estimation can be performed on a subset of alternatives without inducing inconsistency.

Denote the full set of alternatives as F and a subset of alternatives as K. After observing the respondent's choice, we select a set of alternatives K on which the estimation is conducted. Let $q(K|i)$ the probability of subset K to be selected under the researcher's method when choice i is observed.

We assume that for all subsets W not containing alternative i we have $q(W|i) = 0$.

The probability that a person chooses alternative i from the full set is P_{ni}.

The joint probability that the researcher selects subset K and the decision maker chooses alternative i is $P(K,i) = q(K|i)P_{ni} = P_{ni}Q(K)$ where $Q(K) = \sum_{j\varepsilon F} P_{nj}\, q(K|j)$ is the probability of the researcher selecting subset K marginal over all the alternatives that the person could choose.

Therefore we have:

$$\begin{aligned} P_n &= \frac{P_{ni}\, q(K|i)}{\sum_{j\varepsilon F} P_{nj}\, q(K|j)} \\ &= \frac{e^{V_{ni}}\, q(K|i)}{\sum_{j\varepsilon F} P_{nj}\, q(K|j)} \\ &= \frac{e^{V_{ni}}\, q(K|i)}{\sum_{k\varepsilon K} e^{V_{nk}} P_{nj}\, q(K|j)} \end{aligned}$$

when $q(K|j)$ is the same for all $j\varepsilon K$.

This property occurs if, for example, the researcher assigns the same probability for selecting j into the subset when i is chosen and for selecting i into the subset when j is chosen. When this property, named by McFadden(1978) uniform conditioning property, is satisfied, the preceding equation becomes

$$P_n(i|K) = \frac{e^{V_{ni}}}{\sum_{j\varepsilon K} e^{V_{ni}}}$$

which is simply the logit formula for a person who faces the alternative in subset K.

The conditional likelihood function under the uniform conditioning property is

$$_{CLL}(\beta) = \sum_{n} \sum_{i \varepsilon K_n} y_{ni} \ln(\frac{e^{V_{ni}}}{\sum_{j \varepsilon K_n} e^{V_{nj}}})$$

where K_n is the subset selection for n. Maximization of CLL provides a consistent estimator of β. However, since information is excluded from CLL, the estimator based of CLL is not efficient.

In the more general case, when uniform conditioning property does not hold, we have:

$$P_n(i|K) = \frac{e^{V_{ni} + \ln q(K|i)}}{\sum_{j \varepsilon K} e^{V_{nj} + \ln q(K|j)}}$$

In our coding, given the observed choice i, there is but one subset with probability 1 in which the configuration of the chosen product is copied on the other alternatives and a fifth product is added.

Chapter 3

The Hierarchical Logit model

3.1 Introduction

The main problem with the aggregate logit model is that it doesn't allow for respondent's heterogeneity. All respondents are treated the same way and their choices are described by a common set of utilities. In other words, the aggregate logit model is only concerned by what the average people likes. From a marketing perspective, it is a very poor description that misses out market niches and differences between respondents. We know there is great variety in people's tastes and it is a precise interest of the market researcher to know them in all their diversity.

The Hierarchical model is a solution to these issues. It allows each single respondents to have its own tastes - that is, its vector of utilities. Each respondent is considered as a random sample from an underlying population. In marketing studies the respondents are selected to be a representative sample of the whole population, so this is a very realistic assumption.

Since each respondent is a sample from a population, the distribution of such a population is a key feature of the model. Hierarchical model for marketing applications are usually described by the individual-level choice probabilities and the shape of the population distribution. We will first treat the topic in generality and then describe in detail the Hierarchical logit model.

3.2 Hierarchical models for marketing

Suppose that in a marketing survey we have monitored the choices of m respondents (units). Each unit i was the subject of experiments resulting in a vector of observations y_i. For each respondent i we define a vector of parameters θ_i, whose value we want to estimate, representing the specific characteristics of

that respondent. For each respondent i we define the probability of observing a certain vector of choices y as depending only on θ_i as well as some covariates known to the researcher: $p(y_i|\theta_i, x_i)$. The form of this function defines the so called individual-level probabilities. The parameters $\{\theta_i\}$ are called unit level parameters. We consider the parameter θ_i as representing the tastes of a particular respondent that influence the choices observed in the experiments.

We consider each respondent i as arising from an underlying population. We specify a prior distribution of the parameters $\theta_1, \ldots, \theta_m$, describing the variation of tastes in the whole population:

$$\theta_i \sim D(\tau)$$

The parameter τ is usually called the hyperparameter. We also assume that each θ_i is an independent draw from the distribution $D(\tau)$.

We recall Bayes theorem:

$$p(\theta_1 \ldots \theta_m | y_1 \ldots y_m) = \frac{p(y_1 \ldots y_m | \theta_1 \ldots \theta_m) p(\theta_1 \ldots \theta_m | \tau)}{p(y_1 \ldots y_m)}$$

We can write the posterior distribution as:

$$p(\theta_1 \ldots \theta_m | y_1 \ldots y_m) \propto \left[\prod_i p(y_i|\theta_i) \right] \times p(\theta_1, \ldots, \theta_m | \tau)$$

when respondents are independent from each other conditional on unit-level parameters.

The term in brackets is the conditional likelihood and the rightmost term is the joint prior with hyperparameter τ.

Assessment of the joint prior for $(\theta_1 \ldots \theta_m)$ can be difficult due to the high dimension of the parameter space and, therefore, some sort of simplification of the form of the prior is required. One frequently employed simplification is to assume that, conditional on the hyperparameters τ, $(\theta_1, \ldots, \theta_m)$ are a priori independent:

$$p(\theta_1 \ldots \theta_m | y_1 \ldots y_m) \propto \prod_i p(y_i|\theta_i) p(\theta_i|\tau)$$

This means that, given the value of the hyperparameter τ, the behavior of each respondent/unit is independent of the others and that, conditional on τ, inference for each unit can be conducted independently of all other units.

Further more, it is a common choice to assume that the data vector likelihood for a single respondent can be factored in the likelihoods of the single observations.

The specification of the conditionally independent prior can be very important due to the scarcity of data for many of the units. Both the form of the prior and the values of the hyperparameters are important and can have pronounced

effects on the unit-level inferences. For example, it is common to specify a normal prior, $\theta_i \sim N(\bar{\theta}, V_\theta)$ for some fixed $\bar{\theta}, V_\theta$. The normal form of this prior means that influence of the likelihood for each unit may be attenuated for likelihoods centered far away from the prior. That is, the thin tails of the normal distribution diminish the influence of outlying observations.

Use of this sort of normal prior will induce a phenomenon of shrinkage in which the Bayes estimates (posterior means) $\{\tilde{\theta}_i = E[\theta_i|y_i, \tau]\ \}$ will be clustered more closely to the prior mean than the unit-level maximum likelihood estimate $\{\hat{\theta}_i\}$. For diffuse prior settings, the normal form of the prior will be responsible for the shrinkage effects. In particular, outliers will be shrunk dramatically toward the prior mean. For many applications, this is a very desirable feature of the normal form prior. We will shrink the outliers in toward the rest of the parameter estimates and leave the rest pretty much alone.

In general, however, it may be desirable to have the amount of shrinkage induced by the priors driven by information in the data. That is, we should adapt the level of shrinkage to the information in the data regarding the dispersion in $\{\theta_i\}$.

If, for example, we observe that the $\{\theta_i\}$ are tightly distributed about some location or that there is very little information in each unit-level likelihood, then we might want to increase the tightness of the prior so that the shrinkage effects are larger.

This leads to a full Bayes approach in which we specify a second-stage prior on the hyperparameters of the conditional independent prior. This specification is called a hierarchical Bayes model and consists of the unit-level likelihood and two stages of priors:

- likelihood prior $p(y_i|\theta_i)$
- first stage prior $p(\theta_i|\tau)$
- second-stage prior $p(\tau|h)$

The joint posterior for the hierarchical model is given by

$$p(\theta_1 \ldots \theta_m, \tau | y_1 \ldots y_m, h) \propto \left[\prod_i p(y_i|\theta_i) p(\theta_i|\tau) \right] \times p(\tau|h)$$

In the hierarchical model, the prior induced on the unit-level parameters is not an independent prior. The unit-level parameters are conditionally, but not unconditionally, a priori independent

$$p(\theta_1 \ldots \theta_m | h) = \int \prod_i p(\tau|h) p(\theta_i|\tau) d\tau$$

If, for example, the second-stage prior on τ is very diffuse, the marginal priors on the unit-level parameters, θ_i, will be highly dependent as each parameter has a large common component.

The first-stage prior (or random effect distribution) is often taken to be a normal prior. Obviously, the normal distribution is a flexible distribution with easily interpretable parameters.

The use of a hierarchical model for prediction also highlights the distinction between various priors. A hierarchical model assumes that each unit is a draw from a superpopulation or that the units are exchangeable. This means that if we want to make a prediction regarding a new unit we can regard this new unit as drawn from the same population. Without the hierarchical structure, all we know is that this new unit is different and have little guidance as to how to proceed. These assumptions are all very natural from the point of view of marketing practice.

3.3 Inference for hierarchical models

Hierarchical models are ideally suited for MCMC methods. In particular, a Gibbs-style Markov chain can often be constructed by considering the basic two sets of conditionals:

i. $\theta_i|\tau, y_i,\ i = 1 \ldots m$

ii. $\tau|\ \{\ \theta_i\}$

The first set of conditionals exploits the fact that the θ_i are conditionally independent. The second set exploits the fact that $\{\theta_i\}$ are a i.i.d. sample from the distribution $D(\tau)$. Once the $\{\theta_i\}$ are drawn from (i), these serve as data to the inferences regarding τ. If, for example, the first-stage prior is normal, then standard natural conjugate priors can be used, and all draws can be done one-for-one and in logical blocks.

3.4 The Hierarchical Bayes multinomial logit model

This model, as the name suggest, is a hierarchical model based on multinomial logit probabilities at the unit level. We will use the definitions and notations used in chapter 2.

The main difference is that now, instead of a vector of utilities β common to the whole population, each respondent n will have its own vector of utilities β_n.

Let the utility that person n obtains form alternative j in question t be

$$U_{\text{njt}} = \beta_n' x_{\text{njt}} + \varepsilon_{\text{njt}}$$

where ε_{njt} is i.i.d. extreme value and $\beta_n \sim N(b, W)$. As mentioned before, we assume that the vectors of unit level utilities, $\beta_1 \ldots \beta_N$ are i.i.d draws from the normal distribution.

This is equivalent to say that the respondents are a random sample of a population whose utilities are distributed normally with mean b and covariance matrix W.

Giving β_n a normal distribution allows us to speed estimation considerably by using conjugate priors. We will give details on this later on.

We also give priors to b and W. Suppose the prior on b is Normal. If we have no prior information on b, which very likely will be the case, we may use a non-informative prior with unboundedly large variance.

We choose to use an inverted Wishart as the prior for W. If we have no prior information, we may choose as parameters K for the degrees of freedom and I_K as the scale matrix, where K is the dimension of the parameter space and I_K is the K-dimensional scale matrix.

A sample of N respondents is observed. For simplicity of exposition we suppose all people responded to the same number of questions T. The chosen alternatives in all questions for person n are denoted $y'_n = \{y_{n1}, \ldots, y_{\mathrm{nT}}\}$ and the choices of the entire sample are labeled $Y = \{y_1, \ldots, y_n\}$.

The probability of person $n's$ observed choices, conditional on the parameters β_n, b, W is

$$L(y_n|\beta_n) = \prod_i \left(\frac{e^{\beta'_n x_n y_{ni}}}{\sum_j e^{\beta'_n x_{nj}}} \right)$$

The probability not conditional on β_n is the integral of $L(y_n|\beta_n)$ over all possible values of β_n:

$$L(y_n|b, W) = \int L(y_n|\beta)\phi(\beta_n|b, W)d\beta_n$$

where $\phi(\beta_n|b, W)$ is the normal density with mean b and variance W.

The posterior distribution of b, W and $\beta_{1,} \ldots, \beta_N$ is by definition

$$K(b, W, \beta_1 \ldots \beta_N|Y) \propto \prod_n L(y_n|\beta_n)\phi(\beta_n|b, W)k(b, W)$$

where $k(b, W)$ is the prior on b and W described earlier: normal for b times inverted Wishart for W.

Draws from this posterior are obtained through a 3 layers Gibbs sampling. Before describing the layers in detail, we recall three useful Lemmas:

Lemma 3.4.0.1 (conjugate prior for Normal with unknown mean, known variance)

Let the random variable β be distributed normally with unknown mean b and known variance σ. Let $\beta_1 \ldots \beta_N$ an i.i.d. sample of β and suppose the researcher's prior on b is $N(\beta_0, S_0)$.

Then the posterior on b is $N(b_1, S_1)$ where

$$b_1 = \frac{\frac{1}{s_0} b_o + \frac{N}{\sigma}\bar{\beta}}{\frac{1}{s_0} + \frac{N}{\sigma}}$$

and

$$s_1 = \frac{1}{\frac{1}{s_0} + \frac{N}{\sigma}}.$$

Comment: the posterior mean is the weighted average of the sample mean and the prior mean. The weight of the sample rises as sample size size rises, so that for large enough N the prior becomes irrelevant. Also, a larger value of the variance of the prior makes the contribution of the sample more important. As $s_0 \to \infty$ the posterior approaches $N(\bar{\beta}, \sigma/N)$.

The multivariate version of the Lemma is similar.

Lemma 3.4.0.2 (conjugate prior for multivariate Normal with unknown mean, known covariance matrix)

Consider a K-dimensional random vector $\beta \sim N(b, W)$ with known W and unknown b. The researcher observes a sample $\beta_n, n = 1 \ldots N$, whose sample mean is $\bar{\beta}$. The prior on b is multivariate normal $N(b_0, W_0)$. The posterior is again multivariate normal

$$N\left((W_0^{-1} + \mathrm{NW}^{-1})^{-1}(W_0 b_0 + \mathrm{NW}_0^{-1}\,\bar{\beta}), (W_0^{-1} + \mathrm{NW}^{-1})^{-1}\right)$$

If the researcher's prior on b is diffuse (normal with an unboundedly large variance), then the posterior is $N(\bar{\beta}, W/N)$. Taking draws from this posterior is easy. Let L be the Choleski factor of W/N. Draw K iid standard normal deviates, $\eta_i, i = 1, ..., K$, and stack them into a vector $\eta = (\eta_1, \ldots, \eta_K)'$. Calculate $\tilde{b} = \bar{\beta} + L\eta$. The resulting vector $\tilde{b}$ is a draw from $N(\bar{\beta}, W/N)$.

Lemma 3.4.0.3 (conjugate prior for multivariate normal with know mean, unknown covariance matrix)

Let the random variable β be distributed multivariate normally with known mean b and unknown covariance matrix S. Let $\beta_1 \ldots \beta_N$ an i.i.d. sample of β and suppose the researcher's prior on S is $\mathrm{IW}(k, \Psi)$ where k is the degrees of freedom and Ψ the scale matrix.

Then the posterior on S is again Inverse Wishart:

$$\mathrm{IW}\left(N + k, \Psi + \sum_{i=1}^{n}(x_i - b)(x_i - b)'\right)$$

The prior becomes more diffuse with lower k, although k must exceed the dimension of the parameter space K for the prior to integrate to one and have

means. With $\Psi = I$, where I is the K-dimensional identity matrix, the posterior under a diffuse prior becomes $\mathrm{IW}(K+N, (\mathrm{KI}+N\bar{S})/(K+N))$. Conceptually this prior is equivalent to the researcher having a previous sample of K observations whose sample variance was I. As N rises without bound, the influence of the prior on the posterior eventually disappears.

3.4.1 Estimation for the Hierarchical logit model

We choose to use a Gibbs sampler to sample from the posterior. The Gibbs sampler has three layers:

1. $b|W, \beta_1, \ldots, \beta_N$. The β_n's constitute a sample of N realizations from a normal distribution with unknown mean and known covariance matrix. We can therefore use lemma 3.4.0.2. If we are using a diffuse prior, the posterior on b is $N(\bar{\beta}, W/N)$, where $\bar{\beta}$ is the sample mean of the β_n's.

2. $W|b, \beta_1, \ldots, \beta_N$. The β_n's constitute a sample from a normal distribution with known mean b. Given lemma 3.4.0.3 and our prior on W, the posterior on W is inverted Wishart with $K + N$ degrees of freedom and scale matrix $(\mathrm{KI} + \mathrm{NS}_1)/(K + N)$, where S_1 =(1/N)$\sum_n(\beta_n - b)(\beta_n - b)$' is the sample variance of the β_n's around known mean b.SS

3. $\beta_n|b, W$. The posterior for each respondent's β_n, conditional on the observed choices and the population parameters is

$$K(\beta_n|b, W, y_n) \propto L(y_n|\beta_n)\phi(\beta_n|b, W)$$

It is easy to draw from distributions of layers 1 and 2. However, there is no easy way to draw directly from layer 3 so a MH algorithm is used.

For each β_i^n the MH algorithm operates as follows:

a) Start with a value β_n^i

b) Draw K independent values from a standard normal density, and stack the draws in a vector labeled η^1

c) Create a trial value of β_n^{i+1} as $\widetilde{\beta_n}^{i+1} = \beta_n^i + \rho L \eta^1$, where ρ is a scalar specified by the researcher and L is the Choleski factor of W. Note that the proposal distribution is specified to be normal with zero mean and variance $\rho^2 W$

d) Draw a standard uniform variable μ^{i+1}

e) Calculate the ratio

$$F = \frac{L(y_n|\widetilde{\beta_n}^{i+1})\phi(\widetilde{\beta_n}^{i+1}|b, W)}{L(y_n|\beta_n^i)\phi(\beta_n^i|b, W)}$$

f) If $\mu^{i+1} \leqslant F$, accept $\widetilde{\beta_n}^{i+1}$ and let $\beta_n^{i+1} = \widetilde{\beta_n}^{i+1}$.

If $\mu^{i+1} > F$, reject $\widetilde{\beta_n}^{i+1}$ and let $\beta_n^{i+1} = \beta_n^i$.

g) Repeat. For high enough i, β_n^i is a draw from the posterior.

In the MH algorithm, the scalar ρ is specified by the researcher. This scalar determines the size of each jump. Usually, smaller jumps translate into more accepts, and larger jumps result in fewer accepts. However, smaller jumps mean that the MH algorithm takes more iterations to converge and embodies more serial correlation in the draws after convergence. They optimal acceptance rate for the MH algorithm is about 0.44 when K =1 and drops toward 0.23 as K rises. See Gelman et al. (1995, p. 335) for details.

The value of ρ can be set by the researcher to achieve an acceptance rate in this neighborhood, lowering ρ to obtain a higher acceptance rate and raising it to get a lower acceptance rate. In fact, ρ can be adjusted within the iterative process.

The researcher sets the initial value of ρ. In each iteration, a trial β_n is accepted or rejected for each sampled n. If in an iteration, the acceptance rate among the N observations is above a given value (say, 0.33), then ρ is raised. If the acceptance rate is below this value, ρ is lowered. The value of ρ then moves during the iteration process to attain the specified acceptance level.

To sum up, we'll write again in a more concise form the estimation procedure.

We start with any initial values $b^0, W^0, \beta_1^0 \ldots \beta_N^0$. The t-th iteration of the Gibbs sampler consists of these steps:

1. Draw b^t from $N(\bar{\beta}^{t-1}, W^{t-1}/N)$ where $\overline{\beta}^{t-1}$ is the mean of the β_n^{t-1}
2. Draw W^t from $\mathrm{IW}(K+N, (\mathrm{KI}+\mathrm{NS}^{t-1})/(K+N))$, where $S^{t-1} = \Sigma_n(\beta_n^{t-1} - b^t)(\beta_n^{t-1} - b^t)'/N$ ˆ
3. For each n draw β_n^t using one iteration of the MH algorithm previously described, starting from β_n^{t-1} and using the normal density $\phi(\beta_n | b^t, W^t)$

These three steps are repeated for many iterations. The resulting values converge to draws from the joint posterior of b, W, and $\beta_1, \ldots, \beta_N$. Once the converged draws from the posterior are obtained, the mean and standard deviation of the draws can be calculated to obtain estimates and standard errors of the parameters.

The Gibbs sampler converges, with enough iterations, to draws from the joint posterior of all the parameters. The iterations prior to convergence are often called burn-in. Unfortunately, it is not always easy to determine when convergence has been achieved. In practical applications, a sufficiently high number of iterations is performed prior assuming convergence.

During burn-in, the researcher will usually be able to see the draws trending, that is, moving toward the mass of the posterior. After convergence has been achieved, the draws tend to move around (traverse) the posterior. The draws from Gibbs sampling are correlated over iterations even after convergence has been achieved, since each iteration builds on the previous one. This correlation does not prevent the draws from being used for calculating the posterior mean and standard deviation, or other statistics. However, the researcher can reduce the amount of correlation among the draws by using only a portion of the draws that are obtained after convergence. For example, the researcher might retain every tenth draw and discard the others, thereby reducing the correlation among the retained draws by an order of 10. A researcher might therefore specify a total of 20,000 iterations in order to obtain 1000 draws: 10,000 for burn-in and 10,000 after convergence, of which every tenth is retained.

Succinct restatement of the model

Utility:

$$U_{\text{njt}} = \beta_n' x_{\text{njt}} + \varepsilon_{\text{njt}}$$

$$\varepsilon_{\text{njt}} i.i.d. \text{ extreme value}$$

$$\beta_n \sim N(b, W)$$

Observed choice:

$$y_{\text{nt}} = i \text{ if and only if } U_{\text{nit}} > U_{\text{njt}} \forall j \neq i$$

Uninformative priors:

$$k(b, W) = k(b)k(W)$$

where

$$k(b) \sim N(b_0, W_0) \text{ with very large variance}$$

$$k(W) \sim \text{IW}(K, I)$$

Conditional posteriors:

$$K(\beta_n | b, W, y_n) \propto \prod_i \left(\frac{e^{\beta_n' x_n y_{ntt}}}{\sum_j e^{\beta_n' x_{nj}}} \right) \phi(\beta_n | b, W) \forall n$$

$$K(b | W, \beta_1, \ldots, \beta_N) \propto N(\bar{\beta}, W/N) \text{ where } \bar{\beta} = \Sigma_n \beta_n / N$$

$$K(W | b, \beta_1, \ldots, \beta_N) \propto \text{IW}(K + N, \frac{\text{KI} + N\bar{S}}{K + N}) \text{ where } \bar{S} = \Sigma_n (\beta_n - b)(\beta_n - b)'/N$$

Chapter 4

Optional features

Optional features are for us all the extra services or products that can be purchased in addition to a core product.
Examples of optional features include extras in automotive industry, addition to service plans in telecommunication industry, extra warranty coverage in many electronics products and many more.

4.1 Business perspective

From a seller's point of view, offering certain feature as optional has great advantages.
First and foremost, it is a simple way to enrich one's portfolio.
Offering a new feature as an option can make the product appealing to more customers without the risk of alienating the current ones, as the product they know and like is still available to them.

Offering optional features can increase the out-of-pocket expense. Furthermore, optional features can be used to break psychological price barriers. If there is a price barrier at, say, 100 euro, the core product could be offered for 99 euro and a interesting optional feature could be offered for a price way lower than its market value - just a few euro.
At this point, being already over the barrier price of 100 euro, a customer could buy other more expensive options, therefore increasing his/her total out-of-pocket expense.

It is known from marketing literature [1] that the offering of certain options and services makes the product more interesting even for people that are not going to purchase those options.
We will call this Halo Effect. In a rather extreme case, a company may want to offer and advertise an option that almost no customer is actually ever going to

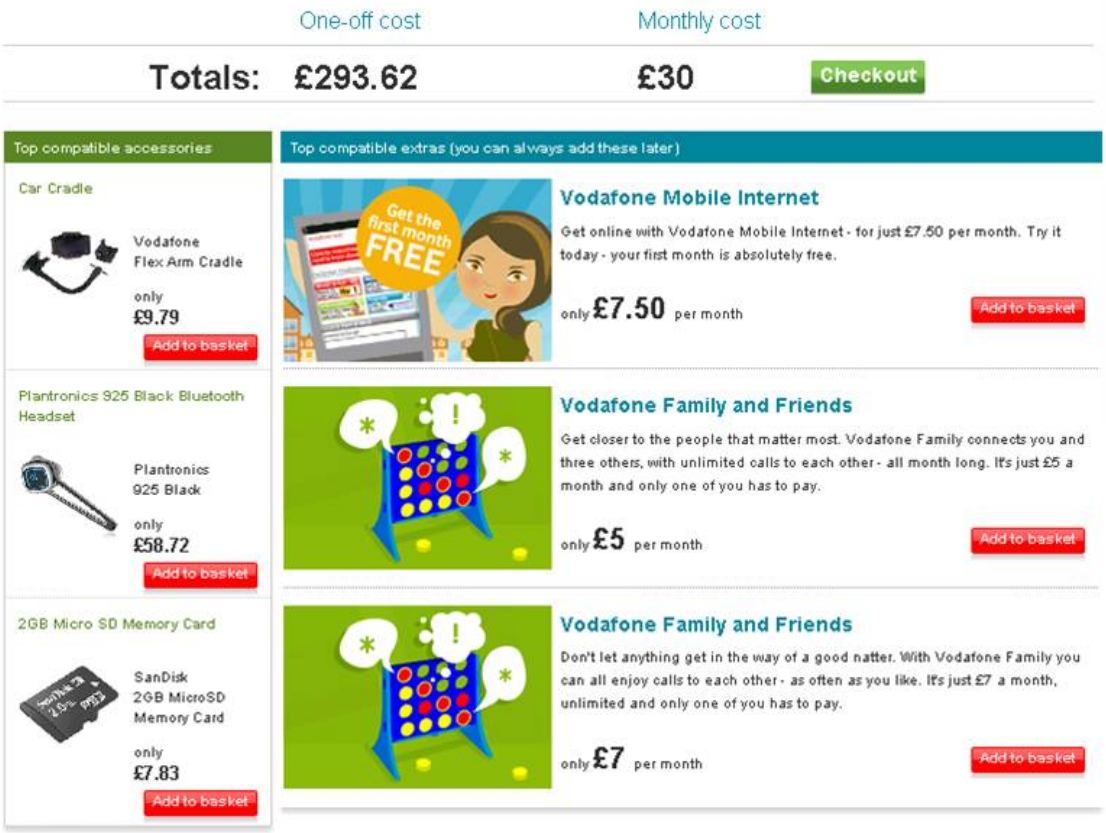


Figure 4.1: Optional features for Vodafone mobile contracts

buy but that makes the product look more interesting.
This may be the case for some advanced technological features: advertising would be focused on the new cutting-edge technological feature, even if very few people will want to pay for it.

The example shown in figure 4 shows other interesting properties of optional features. Customers had the possibility to upgrade their contract any time. It is possible that they gave value to the presence of mobile connectivity, even when not using it right away, because they were thinking of using it in the future.
When it is possible for the customer to buy an optional feature in a second purchase, the simple presence of an optional feature can give value to the product: it gives the customer the opportunity to upgrade at a later stage. This is like saying there is a value in opportunity.

We can see this structure in services where the core feature is usually very cheap (or even free) and that offer the possibility to expand for a price.
For example, VOIP clients like Skype usually offer free client-to-client calls and

chat services.
They offer paid packages for calls to telephone numbers etc. that can be bought at any moment. Therefore, people are mainly convinced to use the client (core product) because of the possibility to expand (call plans).
Also, there is a lock-in effect: when one's contact use a certain client chat system, it is convenient to keep on using the same client.

A completely different example of optional features is the accessories sold with Dell computers.
When buying a Dell, it is possible to customize the specifications of the hardware. This is called Build-your-own or BYO model. Also, it is possible to buy other accessories such as printers, scanners, laptop bags, speakers etc.
All those optional features can be obtained at a later stage from other providers. Since they are usually compatible with all pc's, offering them probably doesn't provide any halo effect. Therefore the main interest of Dell is to have a range of options that customers will buy at the moment of purchase of the core product.

In service markets like insurance, telecom, finance etc. it is generally always possible to upgrade one's contract selecting optional features at a later stage.
An example of optional feature that can't be bought at a second moment is an extension of the warranty period of a product.
We think that the halo effect is sensibly stronger for features that can't be purchased in a second moment, but we didn't investigate in this direction and as far as we know there is no material on this topic.

From the customer's point of view, options are welcome, too.
Options give the sensation of creating a product that is truly fitting one's needs. Also, they can be perceived as a saving: the customer only pays for what he/she really needs and nothing more.

In some extreme case, the choice of options is so wide that a customer can build his/her own product from scratch. This is the case of the BYO (Build your own) products, or mass customization products.
In this case, the customer gets the idea that the resulting product is going to be his/her very own, tailor made according to his/her tastes.

Offering optional features has pitfalls, too.
Offering many options can make a product very customizable to the needs of the buyer, but it can also make the choice decision long and complex.
It is a well established fact, both in practice and literature [1], that optional features can lead to a product being perceived as complicated. This is not only a matter of quantity but also of quality: even when offering a very limited number of options, some of them may take a considerable effort to be evaluated. The result is that the core product is deemed as less attractive: customers may switch to other products that satisfy their needs and are less complicated to evaluate.

A common solution to this problem is to offer only a limited number of optional features bundled in a package. This makes the evaluation easier, but

of course it hinders one of the greatest advantages of optional features: the possibility for the customer to buy just what he/she needs and nothing more.

It is also possible to use a mixed strategy, offering single options and bundles. This is the case of, for example, fast food restaurants. It is possible to buy a menu (a bundle) or just fries and a soft drink (some of the features composing the bundle).
In this case, when customers can see the prices of the single parts of the bundle, they expect a discount on the total price.

4.2 Summary of business questions

The basic business question when dealing with optional feature is *What is the optimal set of options we should offer?*

We can split this very wide question in more precise sub-questions:

- What is the optimal number of options to offer?
- How many people are going to buy a given option?
- Do the options add value to the core product?
- Does the option X make the product complicated?
- Has the option X a halo effect?
- Do people buy certain options together? Should we bundle them?

These are very different questions. We will provide a way to answer most of them by using the powerful tool of choice-based conjoint models.

4.3 Scope of our analysis

We will limit our analysis to products that offer a limited number of options. In other words, we are not going to consider BYO products or other products that can't be easily evaluated at once by the customer. This is because of theoretical and practical reasons related to our choice of the conjoint methodology.
It is a well established fact that conjoint analysis yields good results only if a limited number of attributes per product are shown. When confronted with too complicated products, respondents tends to use simplification strategies and, even worse, get tired more easily providing senseless answers. In other terms, conjoint is appropriate only for studying products that a customer can evaluate in their integrity.

This is not a limitation of this technique. This is because of the way we actually evaluate products: when choosing between a set of products whose features we can actually retain in our minds, we can evaluate their value very

easily. Therefore we use a choosing model similar to the use described by conjoint. When considering extremely complicated products (like BYO products) the main mental task we perform is evaluating the value of a single of those products.

In simpler words, it is very easy to tell between two candy bars which one we like more, because we can have a clear image in our mind of what their features are. When having to select between two offers from different brands of BYO motorcycle from scratch (as Harley Davidson used to offer) we are not able to tell easily which one is better. We actually spend much of our time choosing the options of one of the two companies, instead of deciding which of the two is better. In this sense, forecasting BYO offer we are going to take is not a good question for a conjoint study: most of the mental work is evaluating one of the products, rather than the tradeoffs between one product and the others.

4.4 Methodology

In this section we propose a methodology that is able to answer the business questions put forth in section 4.2. The proposed methodology has been extensively tested in simulations and then applied to data from a real survey.

To find a way to estimate utilities for optional features we start with creating a model for the respondent choice behavior.

The usual conjoint model is based on the fact that customers can keep in mind all relevant aspects of offered products and they can assign a utility to each of them. Their final choice is done by considering at once all the products and selecting one in a way proportional to their utilities. We are going to modify the normal conjoint model to account for optional features.

4.5 Questionnaire structure

Just as in standard conjoint studies, our analysis will be based on questionnaires answers.
We will show a certain number of core products and options to all customers and ask them to choose one core product and some of its options.
In normal conjoint studies what we observe is a choice of a product in a group of n.
In an optional feature study we observe the choice of a product and of some of its options. We don't ask the respondent to state anything about the options present in the other products.
The respondent's task is very similar to an actual purchase. Not asking anything about the other product options we loose a great amount of data but we retain realism in tasks, the most important feature of usual conjoint tasks.

Welk van de onderstaande leasecontracten zou u kiezen?

Aanbieder	TERBERG LEASING	LeasePlan	ALD Automotive	ATHLON CAR LEASE
Dealerkeuze voor aanschaf auto	Vrije keus	Gestuurd	Vrije keus	Gestuurd
Eigen risico	Geen	€ 500	€ 150	€ 500
Prijs basispakket	€ 580	€ 640	€ 630	€ 660
Hulpdienst	Alleen binnenland gratis	Alleen binnenland € 15 ☐	Binnen- en Buitenland € 7,50 ☐	Binnen- en Buitenland + aanvullende extra's* € 7,50 ☐
Vervangend vervoer	Na 24 uur € 15 ☐	Na 24 uur gratis	Na 24 uur € 25 ☐	Per direct € 15 ☐
Fleetreporting	€ 1 ☐		€ 2,50 ☐	gratis
Winterbanden	€ 7 ☐		€ 12,50 ☐	€ 20 ☐
Totaal Prijs	580,00	640,00	630,00	660,00

Figure 4.2: Example of a choice task

4.5.1 Intended result

The result of our methodology is a set of utilities for each respondent.
We want these utilities to reproduce the observed choices of the respondents. By that we mean that, knowing a respondent's estimated utilities we would be able to give the same answers the respondent gave in the conjoint study.

4.6 Assumptions

We argue that offering an option has an effect on the respondent evaluation of the products. This effect is described in [1].

What we want to have is a way to define the utility of the chosen product. We also want a way to define utilities of the other products, those for which we don't observe choices on options. Having that, we can code the optional choice task as a normal conjoint task.

We again assume that respondents have a utility of each feature of a product and the total utility of a product is the sum of the utilities of its parts.
The problem is of course how to define the utility of the options. We also assume that, when selecting an option at a certain price, the utility of that option is added to the core product. When not selecting that option, another utility is added to the core product. This is the utility of "seeing the option and not choosing it".
We assume this is radically different from the utility the respondent would have when not presented with the option at all. That is, offering an option has an

impact of the utility of the product, even for people that don't select it. This is in accordance to marketing theory.

If the chosen product comes with option A at price p1, and option A is selected, the utility of the chosen product will be the utility of its core feature plus the utility of "having A at price p1". If option A is not selected, the utility of the chosen product will be the utility of its core features plus the utility of "not having A".
If option A is not offered in that product, the total utility is the utility of the core product plus utility of "A not offered".

That is, the respondents obtain different utilities for selecting an option at different prices. Of course we can imagine that selecting A at a price of 10 will yield lower utility than selecting A at a price of 5.

On the other hand, the utility that we obtain when not selecting an option does not depend on the price at which said option was presented.
This is a very important concept: we have different utilities for showing an option at different prices, but one utility for not selecting the option, no matter what the price shown.

Also, we assume that people perceive in a different way not being offered an option and being offered an option and deciding not to choose it. In simpler words, the act of offering an option has consequences in the product evaluation, be that option selected or not.
The fact that the respondents sees the option is important. The non chosen option still lingers in the mind of the respondent, therefore changing its perception of the whole product.
In particular, we already know that options can make a product complex or can have a halo effect. We think that, when the utility of the Not Chosen level of an option is higher than the Not Offered level, the option gives a halo effect. When the opposite happens, the option is making the product complicated.

Having defined utilities that respondents get from selecting or not an option, we have to decide how those utilities influence the choice of an option. We considered two models: first choice and share of preference. In the first model, an option is selected with probability 1 if its utility is higher than the utility for not selecting it. In the second model, an option is selected with probability proportional to the exponential of its utility.

4.7 Simulations

We developed our methodology by performing estimation on simulated datasets. To simulate datasets, we generate a questionnaire using Sawtooth SSIWeb. Then we generate utilities for the respondents by sampling from a multivariate normal with given mean and covariance matrix (as described in the conjoint model). We refer to these utilities as "real utilities". We generated three different sets of utilities to use in the simulations. We use the real utilities to generate

answers to the questionnaires using different procedures. In this process we add different kind of random error to further test the robustness of the algorithm.

4.7.1 Answer generation procedures

No Error In the No Error scenario, real utilities are used to generate answers. An option is chosen with probability 1 if its utility is higher than the utility of the Not Chosen option (perfect option choice). The product with the highest utility is chosen with probability 1 (perfect product choice).

Error 1 In the Error1 scenario, options are again chosen basing on the real utilities (perfect option choice). Products are chosen with probabilities proportional to the exponential of their utilities. This is a type I extreme value error, arising from a Gumbel distribution. This is also the error form specified in the conjoint model.

Error 2 In the Error2 scenario, also options are chosen with Gumbel errors.

Using these procedures we generate answers to the questionnaires. As in a conjoint experiment, the answer is the choice of one product and some of its options. In one case, we have an extra procedure that relaxes this assumption:

Perfect choice on all options In this scenario we observe option selection also in the non-chosen product.

This last procedure is mostly of theoretic interest since asking respondents to state anything about non-selected products would destroy the realism of the choice task.
Respondent are very keen on selecting the best product in choice tasks, as they do in life, but would probably give non-realistic results when selecting options of the non-chosen products. They have no interest in evaluating carefully what they are not interested in.

4.7.2 Set of utilities

All set of utilities feature utilities for 300 respondents. For each respondent we generate utilities for 4 attributes each with 7 levels, one of which represents the Not Chosen level. Each respondent answers to 12 fully-randomized questions. These dimensions are similar to the ones of a middle-sized study.

No heterogeneity In this set all respondents have similar utilities. This means that they will tend to select or not to select the same option. This means that, when using perfect option choice, average choice rate for an option level tends to be either 1 or 0. This set was used as a first filter: if an estimation procedure performs poorly on this set it is not suitable for use. The covariance matrix is diagonal, 1*I

Heterogeneity 1-middle heterogeneity In this realistic set respondents have different preferences for options. When using perfect options choice, most

option levels have average choice rates among all respondents that don't collapse to 0 or 1. The covariance matrix is diagonal, 3*I

Heterogeneity 2-high heterogeneity Similar to heterogeneity 1. There is even more variety between respondents. The covariance matrix is diagonal, 5*I

4.8 Estimation procedures

Options coding If option j is chosen in the selected product, we set the value of the same option in the other products as if they were selected.
If option j is not chosen in the selected product, we set its value as Not Chosen and we do the same for the other products.
If an option was Not Present (i.e. for that product it was not possible to select it), it is always coded as Not Present.

Table 4.1: Example of a choice task. The respondent chose product 2, its first option but not the second one.

	Product 1	Product 2	Product 3	Product 4
Core Attribute 1	A1	A2	A1	A2
Core Attribute 2	B1	B2	B2	B1
Core Price (Euro)	10	15	12	20
Option 1	M@10 Euro	M@15 Euro	M@8 Euro	M@15 Euro
Chosen Y/N		✓		
Option 2	N@10 Euro	N@15 Euro	N Not Present	N@10 Euro
Chosen Y/N		×		
Choice		•		

Table 4.2: Options coding: the selection

	Product 1	Product 2	Product 3	Product 4
Core Attribute 1	A1	A2	A1	A2
Core Attribute 2	B1	B2	B2	B1
Core Price (Euro)	10	15	12	20
Option 1	M@10 Euro	M@15 Euro	M@8 Euro	M@15 Euro
Chosen Y/N	✓	✓	✓	✓
Option 2	N@10 Euro	N@15 Euro	N Not Present	N@10 Euro
Chosen Y/N	×	×	×	×
Choice		•		

Table 4.3: Options coding

	Product 1	Product 2	Product 3	Product 4
Core Attribute 1	1	2	1	2
Core Attribute 2	1	2	2	1
Core Price (Euro)	2	4	3	5
Option 1	2	3	1	3
Option 2	5	5	4	5
Choice		•		

Table 4.4: Conversion table

	coded as..
Core Attribute 1	
A1	1
A2	2
Core Attribute 2	
B1	1
B2	2
Core Price (Euro)	
8	1
10	2
12	3
15	4
20	5
Option 1	
M@8 Euro	1
M@10 Euro	2
M@15 Euro	3
M Not Present	4
M Not Chosen	5
Option 2	
N@8 Euro	1
N@10 Euro	2
N@15 Euro	3
N Not Present	4
N Not Chosen	5

Extra product In this case we add a dummy (n+1)-th product to the choice task, a product whose option levels are the same of the chosen product but whose option configuration is the opposite of the chosen product. This is because we want to compare the Not Chosen level against the level at which it was not chosen in the selected product, and to confront the level at which an option was chosen in the selected product against the Not Chosen level.

Table 4.5: Extra product coding

	Product 1	Product 2	Product 3	Product 4	Product 5
Core Attribute 1	1	2	1	2	2
Core Attribute 2	1	2	2	1	2
Core Price (Euro)	2	4	3	5	4
Option 1	2	3	1	3	5
Option 2	5	5	4	5	3
Choice		•			

4.9 Measures of fit

To measure the fit of the model with the data we will mostly look at two kinds of indicators:
-utilities, at a respondent level
-option choice probability, on a aggregate level
In the first case we consider the questions given to the respondent, his/her answers and we see if the respondent's utilities match the information given in the answers.
That is, we see if the utility of an option level that was chosen in the questionnaire is higher than the No Choice utility.
We are not interested in recovering the real utility values but rather to find utilities that reproduce the observed choices. This is the most important measure of fit.
Of course, since each respondent only sees some of the possible options level in his/her questions, for some option levels the information will be borrowed from the whole population.
In the second case we consider each respondent's utilities and we check what option would be chosen basing on these. We do the same with the real utilities and we count the number of times the choice is the same.

This is called Option Count and is calculated over all options taken together.
We also consider the percentage of choice for a single option.
Again, we are not interested in recovering perfectly the observed probabilities but rather to see if the differences are going in a certain direction or if the probabilities collapse to 0 or 1.
It may be the case that the Option Count gives good results but only because all choice probabilities collapse to 1. In this last case, the fit of the model is indeed very poor.
Our aim is not to be able to predict correctly the choice of every single option but rather to have a set of utilities reproducing the observed choices.
In this sense, the value of the prediction choice is not a goal but is anyway pretty important to assess the goodness of the model.
Also, we look at the PCT to see how well the model fit the data. When considering this measure we must consider that adding an extra product increases artificially the PCT, since it is a very unlikely product to be chosen.

4.10 Simulation results

We report the results of the simulations. The tables containing the estimated utilities can be found in the appendix. We advise the readers to look carefully at the questionnaires and the estimated utilities at a respondent's level to consider which estimation method is better.

Scenario name: heterogeneity 1, extra product, no error
Folder: het1 Number of iterations: 20k

Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS 20000 0.169 0.277 0.957 0.933 5.543 8.129

Options prediction percentage: 86/

Exact options choice percentages:

Estimated options choice percentages:

Considerations:

Scenario name: heterogeneity 1, extra product, error 1 Folder: het1E Number of iterations: 20k

Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS

20000 0.132 0.293 0.781 0.703 5.715 5.693

Estimated options choice percentages: 86/

Considerations:

Scenario name: heterogeneity 1, extra product, error 2 Folder: het1E2 Number of iterations: 20k

Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS 20000 0.123 0.337 0.277 0.312 3.473 2.325

Estimated options choice percentages: 54/

Considerations:

Scenario name: heterogeneity 1, extra product, error 2, perfect choice in all options

Folder: het1E2perfchoice Number of iterations: 20k

Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS 20000 0.096 0.313 0.611 0.535 9.960 6.188

Estimated options choice percentages: 85/

Considerations:

Scenario name: heterogeneity 1, extra info, no error Folder: het1Eextrainfo Number of iterations: 20k

Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS 20000 0.102 0.343 0.948 0.925 15.831 7.824

Estimated options choice percentages: 95/

Considerations:

Scenario name: heterogeneity 1, extra info, error 1

Folder: het1Eextrainfo Number of iterations: 20k

Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS 20000 0.104 0.350 0.801 0.736 7.595 5.231

Estimated options choice percentages: 90/

Considerations:

Scenario name: heterogeneity 1, options coding, error 1 Folder: het1nofifth Number of iterations: 20k

Options prediction percentage: 79/Exact options choice percentages:

Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS 20000 0.139 0.350 0.952 0.936 6.620 7.734

Estimated options choice percentages:

Considerations:

Scenario name: heterogeneity 1, perfect choice, no error

Folder: het1perfchoice Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
20000 0.087 0.290 0.895 0.845 7.071 5.138
Estimated options choice percentages: 76/
Considerations:
Scenario name: heterogeneity 2, extra product, no error
Folder: het2 Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
20000 0.125 0.303 0.946 0.917 6.753 6.924
Estimated options choice percentages: 77/
Considerations:
Scenario name: heterogeneity 2, extra info, no error
Folder: het2extrainfo Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
20000 0.129 0.290 0.925 0.898 6.416 6.621
Estimated options choice percentages: 77/
Considerations:
Scenario name: heterogeneity 2, options coding, no error
Folder: het2nofifth Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
20000 0.121 0.300 0.930 0.908 5.821 6.076
Estimated options choice percentages: 66/
Considerations:
Scenario name: heterogeneity 2, perfect choice, no error
Folder: het2perfchoice Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
20000 0.080 0.260 0.901 0.853 7.580 5.010
Estimated options choice percentages: 74/
Considerations:
Scenario name: heterogeneity 1, perfect choice, error 1
Folder: het1Eperfchoice Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
Estimated options choice percentages:
Considerations:
Scenario name: no heterogeneity, extra info, no error
Folder: nohetextrainfo Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
20000 0.169 0.220 0.880 0.841 5.229 7.331
Estimated options choice percentages: 86/
Considerations:
Scenario name: no heterogeneity, options coding, no error
Folder: nohetnofifth Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
Estimated options choice percentages:
Considerations:
Scenario name: no heterogeneity, perfect choice, no error

Folder: het2 Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
20000 0.140 0.307 0.917 0.876 6.457 5.912
Estimated options choice percentages: 59/
Considerations:
Scenario name: no heterogeneity, extra product, no error
Folder: nonhet1 Number of iterations: 20k
Iteration Jump Size Acceptance Pct. Cert. RLH Avg Var Parameter RMS
20000 0.108 0.277 0.972 0.957 71.650 20.724
Estimated options choice percentages: 94/
Considerations:

4.11 Overview of results

Without adding an extra product the model gives quite bad results, especially in the options count of dataset with heterogeneity. This is because the value of the Not Chosen option is always senseless as this option level is never confronted against the other levels. This is the reason why we decided to modify the options coding model to confront the not chosen level against other levels.

When adding the extra product, the model gives acceptable results. The option count and the RLH are sufficiently high; the probabilities of choice tend not to collapse to 0 or 1 although they tend to be quite far from the real ones. In particular, we notice that not all of the information from a single respondent is used.
Observing the utilities on a respondent's level, we see that the model tends to borrow too much from the whole population even when there is valuable information in the answers.
This is especially true when a respondent tends not to choose many levels of a given option.
In this case, each of them appears very few times in the respondent questionnaire.
For example, if a respondent only chooses level 1 of option A, most of his/her coded answers will look something like:

Level A of selected product: NC
Level A of non-selected products: NC, NC, NC
Level A of extra product: A2

when the selected product had a level different from A1 in option A.

Only when level A1 is present in the selected product, the coding will look something like:

Level A of selected product: A1
Level A of non-selected products: A2, A4, A5

Level A of extra product: NC

Let's say that the chosen product features level A1 only once or twice out of 12 questions.

Even if this respondent shows a clear preference of level A1 over the other levels of option A, we are not able to see that from the coding of the questionnaire: level A1 is almost never shown in the same question as levels A2..A5. Also, the latter don't appear as often as they should: most of the time, only once per question!

There is an excessive presence of the Not Chosen level, hence skewing the questionnaire. The algorithm therefore has little information about the non chosen levels and tends to borrow more from the sample. This leads to an overestimation of choice probabilities.

To solve this problem we choose to rebalance each respondent coding by adding extra questions in which levels A1..A5 are shown.
This is a better solution than just setting the parameters to give a better fit at a respondent's level. In case of outliers, their utilities wouldn't be shrunk towards the mean, which is something we want.
We want to borrow less from the sample only for those respondents who are consistent in their answers: therefore we add extra questions only for some of the respondents.
This gives us control on the level of fit we want for each respondent.

The perfect choice model doesn't give a sensibly better fit apart from one case which we will comment later. This is because, just as the base model, this model features the winning product, its opposite and four out of all the possible configurations of the other products.
The fact that results are not so better makes us think that a iterative model in which utilities are estimated many times and option choices in the non chosen product are changed after each estimation will not yield sensibly better results.
In particular the perfect choice model performs poorly in a low heterogeneity scenario. The problem is similar to the one explained before but even more extreme: if many levels of an option are not chosen, now they don't ever appear, not even against the levels of a chosen option. This leads to borrowing a lot from the sample.

We analyzed data collected on options. In the leaseplan study, we were able to see that respondents are extremely consistent with their choices. When selecting an option for a given price in a question, they tend to repeat the same choice in the other questions.
This lead us to the assumption that respondents use a first choice model when choosing options.
If they were using a share of preference model, the choice for a single option level would not be consistent from question to question.

Of course, this may not always be the case. There are studies where the option choice was not always consistent.

We assume that, when inconsistencies arise, they can be explained by price-based interaction.
One such interaction could be a budget constraint: an option level that was selected in the previous questions is not selected now because the price of the core product is already too high.
The existence of budget constraints has been observed in numerous studies. Instead of an actual budget to spend, we can think of this as the maximum amount a respondent wants to spend when buying a product of a certain class.

We will experiment price interactions when studying bundles, in the bundles chapter.

Adding Gumbel errors in the product choice influence only slightly the goodness of fit. The estimation at respondent level is equally good.
This is a very good result, since Gumbel errors are the errors specified in the conjoint model.

Adding Gumbel errors also in the options choice has an extremely negative influence on the goodness of fit. Also, the option count gives very bad results.
Anyway, when price interactions are not present the choice of options tends to be very consistent. This is what we observed in the Optional Features study, the topic of the next chapter.

Since a different price is shown for each single option, we can imagine the customers to be extremely price sensitive and so to always choose the same options.
This lead us to think that type I extreme value errors are not the best to describe the choice of a single option. The strange thing is that in perfect choice model with Gumble errors on products and options, the fit is quite good. No option probability ever collapse to 0 or 1 and their values are actually quite close to the real ones. This may be because, in a high heterogeneity scenario, adding errors in option choice smoothes a lot choice probabilities, avoiding extreme values like 0 or 1. This may explain why the fit is not so great (just a value of about 6) given this great ability to predict options, but it needs more investigation.

In general, the model gives perfect results in the option count when there is little heterogeneity. This is not a good result per se, but is a good test for the model.
In case of more heterogeneity, we can see the option count gives good results and the aggregate choice probabilities, although not numerically correct, don't collapse to 0 or 1.
It seems tough that most probabilities are overestimated.
This is because some levels may appear very few times in a respondent's answers, as they are replaced by the Not Chosen level. This is partially corrected by adding dummy questions for each respondent. These questions don't add external information and they rebalance the questionnaire against the excessive presence of Not Chosen levels.

4.12 Conclusions

The simulation results suggest that the best methodology for no error and error 1 scenarios is the Extra Information methodology.
This methodology leads to utilities that truly match the observed respondent's choices.
All the information given by the respondents is used and the results mirror this fact.

Interestingly, knowing the real option configuration in the other products doesn't help the estimation procedure. Knowing the choice of options in the other products is pretty much all the information we would be able to get from each respondent in an ideal situation. It is good that the Extra Information methodology performs almost as good as if we had this information.

When adding Gumbel errors in the choice of options the results are much worst. We can ask ourselves if adding those errors is really realistic. When analyzing the Leaseplan study data, we were able to see a striking level of consistency in the respondents' behavior. Therefore, we think adding Gumbel errors in the choice of options makes the model unrealistic.

Being satisfied with the simulation results, we proceed to use our methodology on real data.

Chapter 5

Optional features study

This market research study was conducted by SKIM for a car leasing company. The aim of the study was to evaluate and optimize the portfolio of contracts this company was offering.
The study took place in April 2008.
The name of the company requesting the study and other sensitive data in the study have been obscured. The data reprinted here uses fictional names in lieu of the original ones and the names of attribute and levels have been modified.

5.1 Study description

For our purposes, the most interesting part of the study is a series of choice tasks where respondents have to select option.

In each question four products were offered and four options per product were presented.
In each screen the same group of four options was presented for all products. Of course they could differ in level or price.
In total, the four options were chosen from a set of 8 options.

The number of respondents was 229. Each respondent was responsible for the leasing decision in his/her company.
They were divided in three groups according to the fleet size of their company:

- Group 1: 1-24 cars, 110 respondents
- Group 2: 25-99 cars, 71 respondents
- Group 3: 100+ cars, 48 respondents.

The data was collected via web.

Welk van de onderstaande leasecontracten zou u kiezen?

Aanbieder	Brand 1	Brand 2	Brand 3	Brand 4
Dealerkeuze voor aanschaf auto	Vrije keus	Gestuurd	Vrije keus	Gestuurd
Eigen risico	Geen	€ 500	€ 150	€ 500
Prijs basispakket	€ 580	€ 640	€ 630	€ 660
Hulpdienst	Alleen binnenland gratis	Alleen binnenland € 15	Binnen- en Buitenland € 7,50	Binnen- en Buitenland + aanvullende extra's* € 7,50
Vervangend vervoer	Na 24 uur € 15	Na 24 uur gratis	Na 24 uur € 25	Per direct € 15
Fleetreporting	€ 1		€ 2,50	gratis
Winterbanden	€ 7		€ 12,50	€ 20
Totaal Prijs	580,00	640,00	630,00	660,00

Figure 5.1: Screenshot of a choice task with options. The brand names were hidden on request of SKIM.

5.1.1 Description of the attributes and levels

Attributes and levels of the core product:

Brand: seven brands of companies active in the fleet leasing industry. We do not report the name of the brands, as requested by SKIM. **Dealerkeuze voor aanschaf auto:** Vrije keus, Gestuurd.
Gestuurd onderhoud: Ja, Nee, Ja, met haal- en brengservice.
Eigenrisiko: Geen, 150, 500.
Contact leasemaatschappij: Via algemene desk, Via vast aanspreekpunt binnendienst,Via vast aanspreekpunt binnendienst + account manager
Meer/minder km prijs: (minderkm 5 cent, meerkm 7 cent), (minderkm 0 cent, meerkm 10 cent), (minderkm 10 cent, meerkm 10 cent).
Prijs (Euro): 560, 580, 600, 620, 630, 640, 650, 660, 670, 680, 700, 720, 740.

Attributes and levels of the optional features:

Fleetreporting: Niet mogelijk, gratis, Euro 1, Euro 2,50, Euro 5, Not chosen
Winterbanden: Niet mogelijk, gratis, Euro 7, Euro 12,50, Euro 20, Not chosen
KVT regeling: Niet mogelijk, gratis, Euro 1, Euro 2,50, Euro 5, Not chosen.
SVI: Geen, Nationaal gratis, Nationaal Euro 3, Nationaal Euro 6, Nationaal Euro 10, Internationaal gratis, Internationaal Euro 3, Internationaal Euro 6,

Internationaal Euro 10, Nationaal + wasmodule gratis, Nationaal + wasmodule Euro 3, Nationaal + wasmodule Euro 6, Nationaal + wasmodule Euro 10, Internationaal + wasmodule gratis, Internationaal + wasmodule Euro 3, Internationaal + wasmodule Euro 6, Internationaal + wasmodule Euro 10, Not chosen .
Afkoop inleverschade: Niet mogelijk, gratis, Euro 5, Euro 15, Euro 25, Not chosen.
Vervangend vervoer: Geen (geen uitbreiding mogelijk), Per direct gratis, Per direct Euro 15, Per direct Euro 25, Na 24 uur gratis, Na 24 uur Euro 15, Na 24 uur Euro 25, Not chosen
Hulpdienst: Geen (geen hulpdienst mogelijk), Alleen binnenland gratis, Alleen binnenland Euro 7,50, Alleen binnenland Euro 15, Alleen binnenland Euro 25, Binnen- en Buitenland gratis, Binnen- en Buitenland Euro 7,50, Binnen- en Buitenland Euro 15, Binnen- en Buitenland Euro 25, Binnen- en Buitenland + aanvullende extra's, Binnen- en Buitenland + aanvullende extra's, Binnen- en Buitenland + aanvullende extra'sEuro 15, Binnen- en Buitenland + aanvullende extra'sEuro 25, Not chosen.
Brandstofpas: Niet mogelijk, gratis, Euro 1, Euro 5, Euro 10, Not chosen.

5.2 Business questions

The aim of the study was to get an understanding of the lease market.
In particular, they wanted to know what was the best set of options to offer to the public. Also, they were interested in dividing the market in segments.
They also wanted to rethink their core product structure, deciding if it was a good idea to offer as part of the core product features that in the past were only offered as options.

To properly answer all these business questions, it was necessary to find a way to include options in the CBC HB analysis.
Our goal is to find a way to generate utilities that match the observed answers in the questionnaire. We want to match both the product selection and the product selection. Being able to do so, we can use these utilities in market share calculations/scenario simulations.

5.3 Old estimation procedure

SKIM analyzed the collected data using standard conjoint methodologies.
In the questionnaire, respondents had to fill in CBC tasks of the regular kind (no options, just core products) and also some tasks with optional features.

The CBC analysis was carried out on questions without considering the choice of options, that was treated separately. Therefore the results of the esti-

mation procedure was a set of utilities for the attributes of the core product.
The choice of options was analyzed on its own. Counting the most chosen options it was possible to estimate with a good degree of precision what respondents were going to buy.
This simple way of analyzing data gives quite good prediction results, but does not provides utilities for options. This is a major drawback. Imagine the problem is having to choose what options to offer with a product in a certain scenario. For each single person we are able to tell if a certain option would be purchased. What we can't tell is if offering that option makes a difference - that is, if the option is interesting enough to make the respondent shift preferences from a product to another.
To look at the problem from another perspective: since there are no option utilities (and therefore no way to measure complexity increase caused by options) we would conclude that it is better to offer a lot of options. This of course makes no sense from a practical point of view.
Therefore, our new methodology is a great improvement from the old estimation procedure, enabling better and richer scenario simulation.

5.4 The new methodology

Using the insight we gained by using simulated datasets, we applied our new methodology to this dataset.
The best methodology we found was the so called Extra product methodology, which we explained in detail in the last chapter.. Succinctly, its most important characteristics were

- If an option is chosen in the chosen product, mark it as chosen in the other products, and viceversa.
- Add to the choice task an extra product whose option configuration is the opposite of the chosen product

We report the results of the methodology. The choice files and the results of the estimation are in the *optional features study* folder in the attached archive. We advise the interested reader to analyze them carefully. Analyzing the utilities and the answers of each respondent, it is possible to see that the former match the latter.

For each respondent, we calculated if an option would be chosen at a given level. Counting the number of choices of an option at a certain level, we obtain the expected choice frequency of that option.
We can also count the number of times an option was chosen at a certain level in the questionnaire. This is the observed frequency. On a global level, our utilities match with a good degree of precision the observed frequencies.

In the file *option study count* it is possible to see the option count for each respondent. The results are good and consistent among all respondents. On

total, the estimated utilities are able to replicate 84% of the time the exact option choice. This is an excellent result. The fact that we are able to match with such accuracy observed choices gives us great faith in the meaning of our utilities.

This count was based only on the observed answers. The following graph represents expected and measured frequencies compared to the bisectrix. The difference is that, in the second case, we predict what a respondent would do when confronted with an option level he never observed in the questionnaire.
This kind of information doesn't hold much meaning about the methodology. Still, if we want to do scenario simulations, we must care about what happens when we simulate an option choice for a respondent for which he don't have information about the same option. In this case, we want the respondent to answers like the average person in the population.
The algorithm will tend to gather information from the rest of the population - this is enough for normal attributes, but given the particular first-choice structure we used for selecting options, we need to test in detail what happens. The graph shows that the fit seems to be good enough, giving us faith in the results of our scenario simulations.

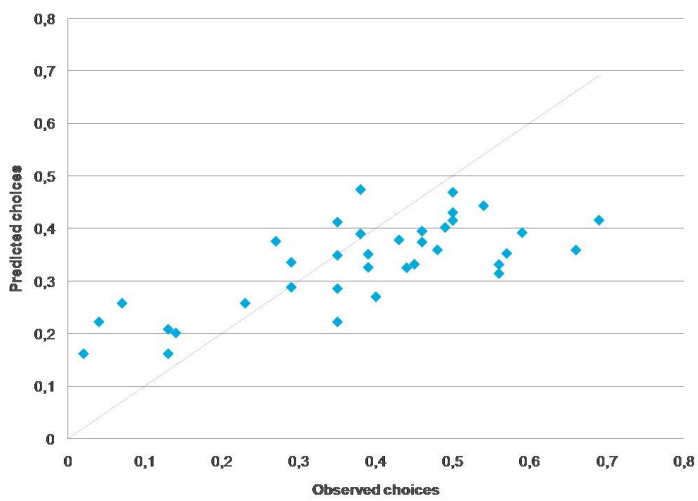


Figure 5.2: Expected and measured choice frequencies for options

5.5 Simulation procedures

To simulate option choice, we used Perfect Option Choice. This methodology was used in the previous chapter to create answers to an optional feature questionnaire.

In Perfect Option Choice, an option is chosen with probability one when the utility of choosing it is higher than for not choosing it.

We have a reason to use this method. Observing the data, we were able to see that respondents were very consistent in their choices of options. If an option was selected at a certain price in one question, it is extremely likely it would be selected again at the same price in the following questions.
We assume that this stems from the particular structure of options: they are inherently simple objects.
An option is made by an attribute and a price. For a person that knows the market, it is very easy to tell if an option is cheap or not, expensive or not. Therefore, evaluation of options tend to be consistent, and selections tend to be consistent.
Of course, there may be cases when this consistency is not present. We think that this can be explained by other reasons, for example by budget constraints logics. If there is a maximum amount of money a person can spend, the selection of an option depends not only on its intrinsic value, but also on the amount that is spent for the core products and the other options. Generally it is considered that, when budget constraints are present, the options with the higher utility per amount of money are selected first.

5.5.1 Improvement over old estimation procedure

First and foremost, the new methodology produces utilities that really match the choices made by respondents. All the information provided by the respondents when answering the questionnaire is taken into consideration and can be retrieved from the coding in the model.

Adding the Not Chosen level it is now possible to forecast choice of options in scenario simulations.
This is a major improvement: it is a utilities-based way to forecast the choice of options.
With this knowledge it is now possible to perform precise forecasts.

It is now possible to know if options add complexity to the product. If the utility for Not Chosen is higher than Not Present, offering the product makes the product complicated. In simple terms, this is like saying that people are happier when they are not offered the option than when they see it (and decide not to buy it).
It is arguable if this interpretation of the results makes sense for all respondents. Does offering an option that I decide to buy increase the complexity of the product? Should we do this analysis only for people that did not/would not select the option?
In a certain sense, it is more prudent to do so. In a deeper sense, adding a lot of very interesting options that most customers will buy makes the product nonetheless more complicated. It does take more time to analyze it, even though the total complexity effect may be diminished by the fact that respondent like the options. What we suggest to do is, for products with a limited number of

options to consider the difference between Not Chosen/Not Present for a given option only for people that would not select that option.

With this methodology it is possible to study the halo effect. If, for a given option, the utility for Not Chosen is higher than for Not Present, it means that people are happier when they see the option (regardless if they are going to buy it or not) than when they don't see it. The interpretation for the halo effect is simpler than for the complexity effect. We can think that the halo effect does not depend on the option selection. It is probably even stronger for the people that are buying the option.

In conclusion, with the new methodology a whole new range of business questions can be answered.
Now it is possible to analyze the effect each option has on the global evaluation of the product. This is extremely important in analyzing a scenario. We are able to tell if offering an option will move customers from one product to another. This can happen in two ways: the option may be increase interest for customers, or drive them away by making the product too complex. This information is extremely important when taking strategic decisions.

Chapter 6

Bundles

Bundles are defined as a collection of features or services that are offered together, for a single price, as an addition to some product or service.
They are closely related to optional features and have similar advantages and disadvantages but their structure is much more complex to analyze.

From the point of view of the consumer, the existence of bundles is important for two characteristics:

- convenience: when a consumer has to analyze complex offering of products that are meant to work together (es. Selecting a fully comprehensive financial service that covers funding, pension, insurance etc), an offer featuring a sensible combination of many products provides a solution. A consumer with limited expertise in the market may be interested in making a single choice instead of many difficult one.

- discount: it is very common for the bundle features to be also sold on their own. Usually when buying a bundle of features the price of the bundle is lower than the total price of the features considered singularly. A customer interested only in some features may buy a bundle containing other not so interesting features if the total price is low enough. This behavior is well established in the experimental literature.

From the seller's point of view, the reason for offering bundles is mainly to increase the out of pocket expense. When a must-have feature is only offered in a bundle, the customer is forced to buy the whole bundle, paying for features that may be uninteresting for him/her.
This may not be seen as unfair by the customer - there is evidence in literature ([11], [2]) that customers accept as fair a bundle with some uninteresting features as long as all the components in the bundle "make sense" when considered together.
For example, in a house insurance product, a service bundle containing extra insurance for flooding, fire and structural failure may be considered fair as the three features are pertaining to the same area.

When features are offered in bundles or on their own, a customer may want to buy a bundle even if it contains uninteresting features, as long as the total price is low enough. Therefore, offering a bundle increases the total out of pocket cost.
This behavior is well known and it has been observed many times.

6.1 the Bundle study

SKIM was much interested in finding a way to analyze bundles in a CBC context. It was decided to test methodologies based on optional features on simulated datasets and, if the results looked promising, test them on real data.
Tests were run on simulated datasets and the results looked good. We therefore decided to try them on real data. The company had no datasets about bundles arising from previous studies, so it was decided to organize a field study just to test these methodologies.

The description of this study and the estimation results are found in the next chapter. In this chapter we'll describe assumptions and considerations that draw our analysis.

6.2 Differences and similarities with optional features

In our classification, we divide all bundles in two possible types.

We can discriminate between bundles that offer services that can be substituted or not. In the first case, one could buy a product without the bundle and obtain in a second moment products equivalent to the one of the bundle. The provider in this second purchase could be the same or a different company from which he/she bought the product: in this case we are speaking about (delayed) upgrades.
When a service can be bought only from the provider of the core product and only at the moment of purchase, the seller has considerable power when the bundle has some "must have" features.

We limit our investigations to bundles that can only be provided by the company providing the core product. This makes the analysis much easier.

We will not consider in detail the case of delayed upgrades. We will not investigate the likeliness of upgrading at a second stage. We only want to know which bundles are going to be selected at the moment of purchase of the core product.

Our main assumption about bundles is that there is such a thing as a bundle price elasticity. This means that it is possible to study the price elasticity of a bundle without knowing what items it has. This is clearly an oversimplification: two bundles with the same price and different items clearly have different

elasticities.
This is especially evident if one thinks of the case of single-item bundles.

Still, this simplification works pretty well with bundles made of many items. Also, from a purely practical point of view, studying the bundle price elasticity as depending on the bundle content would mean finding a way to divide the total price among bundle components. This would of course give better results. Still, it would be very complicated. It is our opinion is that it is not possible to use such an approach using only the information gathered from a conjoint-style study.

6.2.1 Interaction effect

In many real life examples bundles are often rather expensive compared to the core product and can constitute a significant fraction of the total price.
Therefore, the interaction between core price and bundle price can be very strong. We did not study in detail this interaction when modeling options since we decided to consider only options whose price was little compared to the core product price.
If budget constraints behavior happen in the choice of options, they are easily captured in the heterogeneity of the model. Also, offering many options for limited prices, it is easy to see if psychologic threshold are present. If customer are for example shy of spending more than 100 euro but they have enough options to arrive close to that amount it is quite easy to detect this threshold as many will buy options for a total just under 100 and other will buy option for a total well over 100, leaving a space between 100 and, say 120, with little observed choices.
When studying bundles it is harder to detect such behavior, since bundles may constitute a huge part of the total price.

When designing the bundle study, we decided to study a class of products that had a strong interaction between bundle and core price.
Our choice was the market of credit cards.

6.2.2 Goals of the model

The goal of our bundle model is to provide utilities on a respondent level.
The data used to estimate utilities must only come from realistic choice tasks.
These utilities describe core products and bundles and reproduce the observed answers. By this we mean that only knowing the utilities we would be able to answer the questionnaire in a way similar to the one of the respondents.

6.2.3 Limitations of the model

As we will describe in detail in the next chapter, we studied the problem of bundles when a single bundle is offered with each product.
In reality, the situation can be much more complex. For example, with a single

core product many extras may be offered at once. For example, one might have many optional features and bundles made of some of those optional features.
Our simple model is not meant to analyze such a situation. One could use it to analyze complex situations - after all, knowing utilities one can simulate choices of any combination of bundles and options. Still, we advise not to take the analysis too far: given a complex situation, with many interrelated bundles and options, unknown dynamics might be present in the choice procedure.

6.3 Utility from a business perspective

As we will show in the next chapter, we were able to create a working methodology for bundles.
From a business point of view, a bundle methodology leads to great answers.
The result of the estimation is a set of utilities for each respondent.
With this utility we can simulate the respondent's behavior in various scenarios.
That is, we can perform realistic simulation.

We therefore can make prediction on the number of people that will select a certain bundle, we can assess market shares, income forecasts and cannibalization risks.

Also, these results can be used in product definition: we can find the optimal bundle for a certain objective (es increasing market share, increasing revenues etc).

Chapter 7

Results of the Bundles study

Results of the Bundle study

7.0.1 Characteristics of the study

The Bundle study is a study devised to test new methodologies. The goal is to find a way to handle bundle data arising from conjoint-style choice exercises.
This study was funded and managed by SKIM Analytical and its results are published by kind permission.
The topic of the study is credit cards.

The study was directed to Dutch respondents. It was administered via internet until a total of 250 completed entries was reached.
Data was collected from the 23th to the 28th of July 2009.

The questionnaire started with questions about how credit card are used: how often, to buy what and how much their yearly fee is.
Then the respondents were requested to perform 12 choice tasks, similar to the ones used in conjoint studies. They only difference is that for each product shown on screen it was possible to select a bundle of services for an extra price.
The questionnaire finished with more questions about demographical details and likeliness of choosing a new credit card in the next future.

There was no particular screening to select the population of respondents. People of all ages were invited to take part in the questionnaire. However, their answers to some questions could stop them from completing the questionnaire. Most notably, people who had a credit card but used it too few times in a year were screened out as considered not interesting for the questionnaire.

Structure of the questionnaire

1. How old are you?
1. Less than 18 years old - **disqualified**
2. 18-29 years old
3. 30-39 years old
4. 40-49 years old
5. 50-64 years old
6. 65+ years old -**disqualified**

2. Do you have a credit card?
1. Yes, one that is my own and one from my employer
2. Yes, only from my employer
3. Yes, only my own
4. No, I don't have a credit card

3. What is the brand of your credit card? If you have more than one, choose the one you use more often
Pictures of 34 credit cards from ABN AMRO, American Express, ANWB, Fortis, ING, Postbank, Rabobank, SNS, Visa appear.

4. How often do you use your credit card?
1. Once per year or less - **disqualified**
2. Once every three months
3. At least once per month
4. Each week
5. Many times a week
6. Each day

5. Who in your household decides to purchase or to modify the contract (cancel, renew, change, etc.) of your credit card?
1. Me
2. My partner - disqualified
3. Me and my partner

6.Do you have more than one (personal) credit card?
1. Yes
2. No

7. From what bank or institution is your private credit card?
Reporting the selection in question 3

8. Do you know what is the yearly fee of your personal, most used credit card?
1. Yes, it is Euro ... per year

2. No

8b. If not, what range is it in?
1. From 0 to 5 Euro per year
2. From 5 to 15 Euro per year
3. From 15 to 25 Euro per year
4. From 25 to 35 Euro per year
5. From 35 to 45 Euro per year
6. From 45 to 55 Euro per year
7. More than 55 Euro per year
8. I don't know

9. What do you use your credit card for? Multiple answers possible.
1. Buying gas
2. Booking holidays (either by travel agency or by internet)
3. Buying on the internet (no holidays)
4. Groceries shopping (total cost of less than 100 Euro)
5. Groceries shopping (total cost of more than 100 Euro)
6. Going out
7. Other activities, namely

10. Divide in percentage all your credit card purchases among the following categories:
List of answers chosen in the previous question

11. Please specify the percentages of your transactions taking place in the Netherlands and in foreign countries

12. How likely are you to change your credit card in the next two years?
Very likely
Likely
A bit likely
I don't know
A bit unlikely
Unlikely
Very unlikely

12b. If unlikely: what is the main reason you find it unlikely to happen?

12c. If likely: what is the most important reason you find it likely to happen?

13. Please indicate how much do you agree with each of the following statements

A list of statements appears. For each statement it is possible to select one among: *Totally agree, Agree In part, Don't agree nor disagree, Disagree in part, Disagree, No opinion*

The statements are:
Each credit card brings with it a bit of insecurity
With a credit card I never need to worry that I don't have money
With a credit card I find myself in debt easier
My credit card provider to protect me protection against building of a big debt
A credit card gives me status
A credit card is practical
I'm afraid that somebody could commit a fraud with my credit card
A credit card gives me services that not everyone has

14. Choice tasks

Twelve choice tasks, each featuring 4 products, each one with one bundle, and a None of the above/No choice option.

15. What is your sex?

1. Male
2. Female

16. Which of the following better describes you?

1. Living with my family
2. Living on my own
3. Married/living with my partner
4. Student house/residential community
5. Other, namely

17. How many children are there in your house?

18. What is your income range?

1. Less than 20.000 Euro/year
2. Between 20.000 and 30.000 Euro/year
3. Between 30.000 and 40.000 Euro/year
4. Between 40.000 and 50.000 Euro/year
5. Between 50.000 and 60.000 Euro/year
6. Between 60.000 and 70.000 Euro/year
7. Between 70.000 and 80.000 Euro/year
8. More than 80.000 Euro/year
9. I prefer not to answer this question

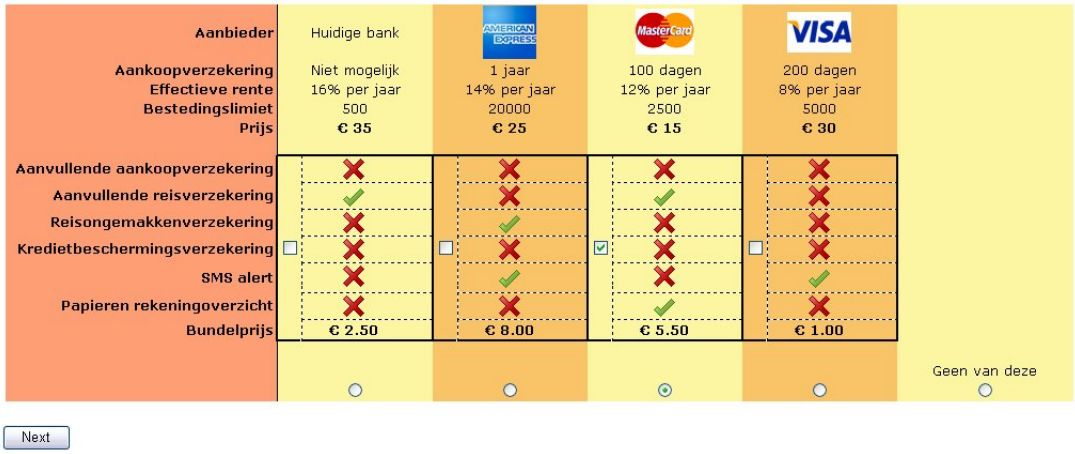
19. What is your higher level of education? 1. Lagere school / none
2. MAVO / MULO / lager beroepsonderwijs
3. HAVO / VWO / HBS
4. Middelbaar beroepsonderwijs
5. Hoger beroepsonderwijs
6. University

20 Which of the following options better describes your working situation?
1. Full-time employed worker
2. Part-time employed worker
3. Full time independent worker
4. Part-time independent worker
5. I am a housewife/househusband
6. I don't work
7. I am a student
8. I am retired
9. No answer

21. What are the four cyphers of your postcodes?

7.0.2 Choice task design

For each respondent 12 questions were generated. Each question featured 4 products, each coming with a bundle.
The respondent could choose one of the 4 products or select a No Choice option.
This picture shows the choice task structure:



7.0.3 Attributes

The core product was defined by the following attributes and levels:

- **Brand:** ING Rabobank, SNS Bank,, ABN AMRO, Fortis, VISA, Mastercard, American Express
- **Price:** Euro 10, Euro 15 , Euro 20, Euro 25, Euro 30, Euro 35, Euro 40, Euro 45, Euro 50, Euro 60, Euro 70
- **Effective interest:** 8% per year, 10% per year, 12% per year, 14% per year, 16% per year
- **Limit on CC:** Euro 500 Euro 1.500, Euro 2.500, Euro 5.000, Euro 10.000, Euro 20.000
- **Purchase insurance:** Not present, 100 days, 200 days, 1 year

The bundles were defined by the following attribute and levels:

- **Additional purchase insurance:** 100 extra days, 200 extra days, Not present
- **Additional travel insurance:** Present, Not present
- **Travel inconvenience insurance:** Present, Not present
- **Credit protection insurance:** Present, Not present
- **SMS alert:** Present, Not present
- **Statement fee:** Present, Not present

Level balance

The questionnaire was generated in two separate moments.
First, the part relative to core products and then the part relative to bundles was generated.
Core products were generated automatically by using Sawtooth SSI Web software. A balanced design (showing levels of an attribute with equal frequency) was chosen.

A similar feat was obtained for bundle levels.
We decided to study bundles with one to four items.
On a global level, we wanted bundles to have this distribution: 20% one item, 30% two items, 30% three items and 20% four items.
Therefore the coding is not perpendicular: the Not Present level will be seen more often.

We also wanted to achieve balance on a respondent's level.
On the 48 bundles each respondent sees (4 bundles for each of the 12 questions), we want the global percentages to be observed. Therefore we decided to allocate for each respondent 10 bundles with one and three items and 14 bundles with two and four.
We also wanted balance on a respondent level, i.e. we want each respondent to see a given level the same number of time.
This is the way we created the bundles for each respondent.
First we allocated for each respondent a number of levels values from which to draw.
For example, for Attribute 1, we can draw level Not Present 28 times and Present 20 times. The distribution is such that, for each respondent, it is possible to arrange in some way the levels to form 10 bundles with 1 feature, 14 with two etc.
We create bundles randomly sampling without replacements from this distribution.
We sample a level from each attribute and consider the resulting bundle. Then we see if the bundle is not suitable: for example, if it has 6 items, or if it has 3 and we already created all 3-items bundles for that respondent.
If this is the case, we "put back" the drawn levels and draw another bundle. If we have to repeat a draw more than a certain number of times, we conclude we are in a dead end and start all over again for that respondent.

This is not the optimal way to create a questionnaire but it is very fast. The VB code to generate the questionnaire is contained in the *bundle gen.xls* file.

Bundles generated with the VB code and core products generated by SSI Web were mixed randomly to create questions. From a practical point of view, this means that expensive bundles can be offered with cheap credit cards and viceversa.

Discount structure

The price of bundles was calculated on the following way.
For each item present in the bundle, a price was generated. The price for each item is random from a list of values. These values are:

- **Additional purchase insurance:** 1 Euro, 2,5 Euro, 5 Euro
- **Additional travel insurance:** 1 Euro, 2,5 Euro, 5 Euro
- **Travel inconvenience insurance:** 2,5 Euro, 5 Euro, 7,5 Euro
- **Credit protection insurance:** 2,5 Euro, 5 Euro, 7,5 Euro
- **SMS alert:** 1 Euro, 2,5 Euro, 5 Euro
- **Statement fee:** 2,5 Euro, 5 Euro, 7,5 Euro

The price of the bundle is then the sum of the singular items price, with a discount. If the bundle has two items, the discount is 10 % , if it has three the discount is 20 % and it is 30 % when there are four items.

A discount structure is very realistic feature, since bundles usually have a lower total price than single goods. This is of course to stimulate a higher out-of-pocket expense.

Total price

During the design phase it was a matter of discussion wether respondents should see on screen the total price (core+bundle) when they decide to select a bundle. It was decided that this price should not be shown automatically on screen. Showing it would probably enforce a stronger interaction between core price and bundle price and it wouldn't add realism to the questionnaire.

Filtering

The only post-hoc filtering of respondents was based on their answering time. The average time to answer the whole questionnaire was estimated as 15 minutes. Respondents finishing the questionnaire in less than 5 minutes were taken out of the sample.
This resulted in a sample of 205 respondents from the original 250.

7.0.4 The choice tasks

After the filtering, a bundle was selected in 194 choice tasks out of 1709 (11,35 %). This percentage is calculated only considering the questions whose answer was different from the None/no choice answer.
Only 49 respondents out of 205 selected a bundle at least once.

The No choice answer was selected 751 times out of the 2460 questions (12 questions * 205 respondents), resulting in a percentage of 30,5 % .
This is a pretty high percentage. The number of No choice selections is not increasing question after question (from question 1 to 12: 52 60 61 60 65 63 68 63 64 69 59 67).
This makes us think that these answers are to be taken as serious.
Probably respondents had a very precise idea of what kind of card they look for and they are not ready to settle for something else. We studied a rather large range of prices for credit cards (from 10 Euro to 35 Euro yearly fee) so it is more probable for one respondent not to have a card in his/her favorite price range among the 4 on screen.

We tried to see if bundles were selected more often by a certain class of respondents. In particular, we wanted to know if there was a link between the current yearly expenditures and bundle selection.

Table 7.1: Distribution of respondents selecting at least one bundle

expenditure range	respondents that selected bundles	number of respondents in class5cm	percentage
0-5 euro/year	3	15	20,00%
5-15 euro/year	6	30	20,00%
15-25 euro/year	11	44	25,00%
25-35 euro/year	8	27	29,63%
35-45 euro/year	5	18	27,78%
45-55 euro/year	4	9	44,44%
more than 55 euro/year	1	8	12,50%
don't knot/don't say	11	53	20,75%
	49		

What we can see from the table is that even respondents in the low paying range selected bundles.
It seems that bundle selection gets slightly more frequent as yearly fee rises. This increase is interesting, but the discovery that also low and middle range spenders select bundles has far more important commercial consequences.

Coding for the CBC-HB estimation

To run the CBC-HB algorithm on the data, the bundles were coded in the following way:

- If a bundle was selected, the attributes of the bundles of all products in the choice task are left untouched. An extra product is added. Its core product levels are the same of the chosen product. All its bundles levels are coded as Not Present, and it's price is coded as 0, meaning it doesn't have an utility that is taken into considerations in the algorithm.

- If no bundle was selected, the bundle is coded with the new value "Not present" for all the products in the choice task. The bundle price for all products is 0, meaning that an utility of 0 is used in the iteration. An extra product is added. This product is equal to the chosen product but its bundle has the original attribute values of the chosen product, price included.

a picture will make it clearer.

The coding is therefore:

This coding is different from the coding used for options for two main reasons:

- Price: the price of non-chosen options was used as data in the estimation, the price of non-chosen bundles is not

Table 7.2: Answer of a bundle choice task (chosen bundle)

	Product 1	Product 2	Product 3	Product 4
Core Attribute 1	A1	A2	A1	A2
Core Attribute 2	B1	B2	B2	B1
Core Price (Euro)	10	15	12	20
Bundle attribute 1	Present	Not Present	Present	Not Present
Bundle attribute 2	Not Present	Present	Present	Present
Bundle Price (Euro)	4	5	3	4
Chosen Y/N		✓		
Choice		•		

Table 7.3: Coding (chosen bundle)

	Product 1	Product 2	Product 3	Product 4	Product 5
Core Attribute 1	1	2	1	2	2
Core Attribute 2	1	2	2	1	2
Core Price (Euro)	1	3	2	4	2
Bundle attribute 1	1	2	1	2	2
Bundle attribute 2	2	1	1	1	2
Bundle Price (Euro)	2	3	1	2	0
Choice		•			

Table 7.4: Answer of a bundle choice task (bundle not chosen)

	Product 1	Product 2	Product 3	Product 4
Core Attribute 1	A1	A2	A1	A2
Core Attribute 2	B1	B2	B2	B1
Core Price (Euro)	10	15	12	20
Bundle attribute 1	Present	Not Present	Present	Not Present
Bundle attribute 2	Not Present	Present	Present	Present
Bundle Price (Euro)	4	5	3	4
Chosen Y/N		×		
Choice		•		

Table 7.5: Coding (bundle not chosen)

	Product 1	Product 2	Product 3	Product 4	Product 5
Core Attribute 1	1	2	1	2	2
Core Attribute 2	1	2	2	1	2
Core Price (Euro)	1	3	2	4	2
Bundle attribute 1	2	2	2	2	2
Bundle attribute 2	2	2	2	2	1
Bundle Price (Euro)	0	0	0	0	3
Choice		•			

Table 7.6: Conversion table for attributes

Core Attribute 1	
A1	1
A2	2
Core Attribute 2	
B1	1
B2	2
Core Price (Euro)	
8	1
10	2
12	3
15	4
20	5
Bundle attribute 1	
Present	1
Not Present	2
Bundle attribute 1	
Present	1
Not Present	2
Bundle Price (Euro)	
3	1
4	2
5	3

- Not Present & Not Chosen: for options there was a difference when an option was shown but not selected or when it was not selected at all. For bundles only the Not Present level is used.

This coding was developed basing on the ones used for options, with Not Present and Not Chosen levels, but the prediction rate was low.
Also, for respondents that selected many bundles, the two levels tended to have similar values.
Other codings we tried, and at the end the best results were obtained by overriding the Not Chosen attribute with Not Present.

7.0.5 Results

With this coding we are able to predict $69,38\%$ of the bundles. This is an excellent result, meaning the prediction rate is good enough to be used in practice.
This percentage only refers to bundle prediction: given the utility and the bundle features alone, the number of time the predicted bundle choice is the same as the observed one.
For the alternatives prediction rate (that is, prediction of which of the four product will be chosen) the result is equally good with a rate of over 70%.
These results were obtained with constraints on the price (utility of lower prices $>$ utility of higher prices).

Table 7.7: Prediction rates

	Observed selections	**Predicted selections**
Question 1	20	10
Question 2	21	11
Question 3	25	17
Question 4	21	12
Question 5	19	9
Question 6	19	9
Question 7	18	6
Question 8	17	11
Question 9	14	5
Question 10	13	7
Question 11	18	7
Question 12	13	4

Looking at the prediction rate it is possible to see where it is possible to improve. The distribution clearly underestimate bundle selections.
Looking at the single predictions as reported in *bundle results.xlsm*, sheet *bundle counter 2*, it is possible to see that the algorithm creates both false positives and false negatives. This is a good feature, meaning that the algorithm doesn't

"need" an overwhelming amount of recorded choices to decide to predict a bundle selection. This was actually the problem with the first, optional features-like coding.

7.0.6 Comment on the coding

The coding of bundles is not just a trick to improve estimation, it has some important consequences.

Its differences from the optional features coding tell us a lot about the way bundles work.
For example, the Not Chosen level is not present anymore in this coding. This level was used for optional features to indicate an option that was offered and not chosen, while Not Present indicated an option that was not featured and could not be chosen.
For bundles, the two are the same. For options, the difference in utility between Not Chosen and Not Present meant that some options added value to the core product even when not selected. In many cases, using the same coding for bundles, the difference between the two values was very low and the prediction rate was not satisfactory.
Even if the two parameters are, in a certain way, the same, writing it as two makes the design less orthogonal and therefore the estimation not as precise.
This difference means that bundles are considered as one and, if not selected, their presence has no influence on the evaluation of the core product.

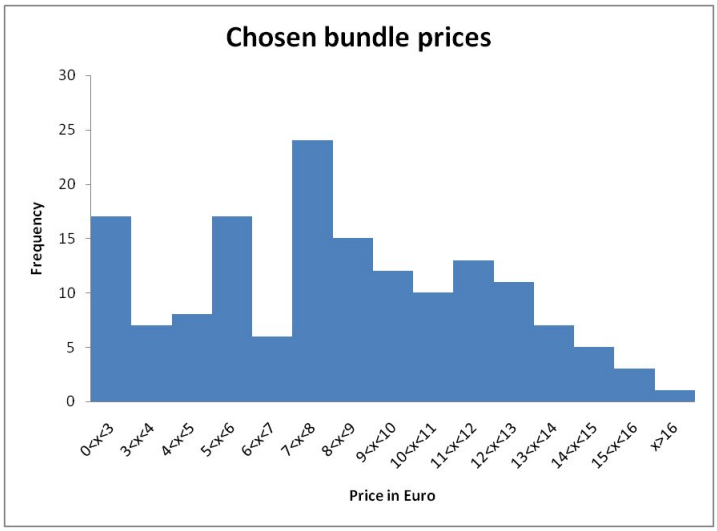
Speaking of prices, for options the price of an unselected option was taken in consideration much more heavily.
If an option was selected for a certain product, the prices of the same options in other products were included in the coding.
Also, when an option was not selected it was recorded the price at which it was not chosen. In our interpretation, this meant that the price at which an option was offered was important. The information was processed and kept in mind by the respondent.
For each option the price was a part of the levels: for example for option A possible levels would be: Not Present, Not Chosen, Present at 1 Euro, Present at 2 Euro etc.
This gave us the possibility to measure the price sensitivity for each option.

Now, for bundles, the situation is different. It is not possible to define price sensitivity for single bundle items if we don't know how respondents allocate the total bundle price between items.
Therefore, it is only possible to measure price sensitivity for the whole bundle.
This could be due to the fact that bundles are inherently more complex than options.
People processes their decision on a bundle and then discard the data, while the price of options - simpler objects- lingers on in the respondent's mind. This causes an halo effect.

It seems that for bundle there is no halo effect. Bundles are important only if selected. If not, it doesn't matter what items they were made of, or at what price were they offered.
Therefore, bundles should only be used to increase out-of-pocket expense, knowing that no halo effect will come from them.
This result can bring give guidance in strategic decisions. For example, if one firm has an option that has some positive halo effect, they should not put it in a bundle. On the other hand, since bundles are considered at once and suffer less from the complexity problems, it may be a good idea to put many complex options in a bundle.

Bundle analysis

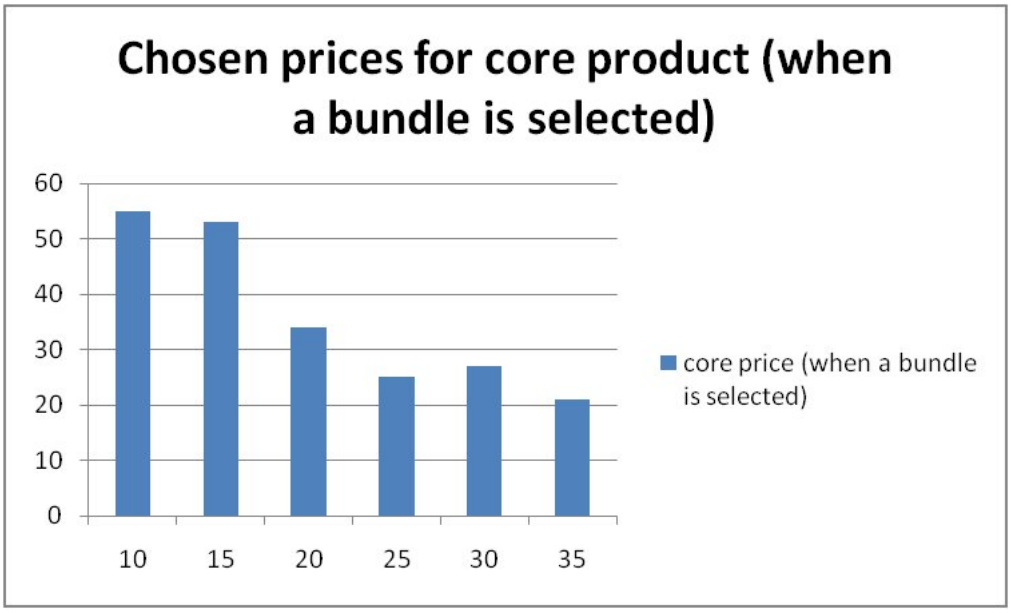
We plot on a graph the distribution of prices for chosen bundles:



We can see that the distribution is clearly non-symmetrical.
This is no surprise, as we imagine cheaper bundles to be more interesting. There seems to be a decrease between the cheapest bundles and the middle priced ones. This may be due to two reasons.
This may be due to the inherent structure of bundles: given the way bundles are priced, it is less likely for a bundle with the expensive features to land in the area between 3 and 5.
Also, we expect uninterested people to buy bundles if cheap enough.
Even if people are not really interested in the bundle content, if the price is so low to be perceived as irrelevant on the total credit card purchase price, they are likely to buy the bundle.

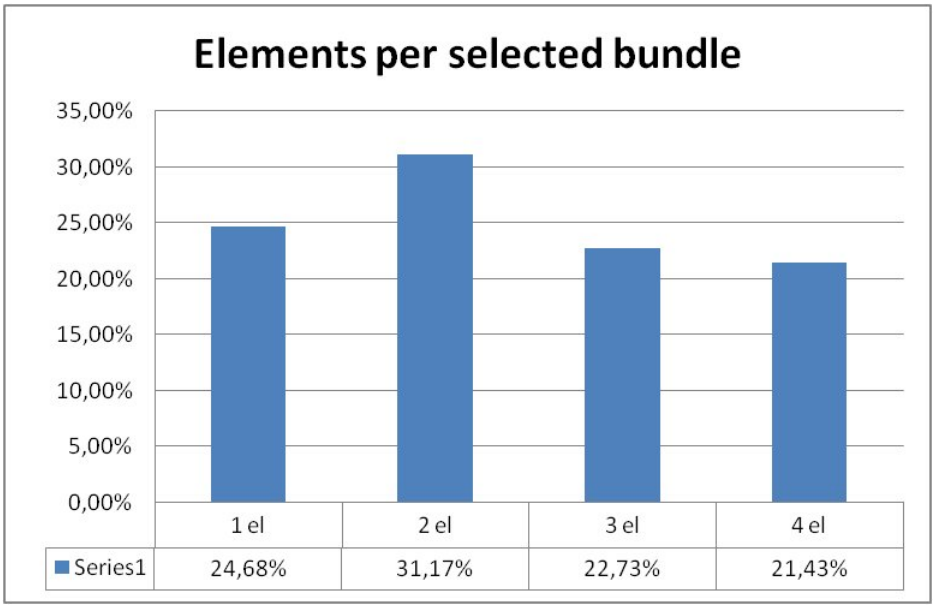
The average price for a bundle is 7.82, coincident with the highest bin in the histogram.

The distribution of prices for the core product is different. We can see it is way more regular and people have a clearer predilection for cheaper product, as it is in most markets.
Looking again at the bundle graph, one may think that this is not happening for bundles and that people are more interested in more expensive products.
This is actually what also happens in bundles: most of the people selected the cheapest product, that is they decided to pay zero and select no bundle. To make the graph complete, there would be a column indicating the zero-selection, and it would be more than 10 times the height of the tallest column.



We can also notice that, when considering only the core prices when a bundle was chosen, the distribution is not at all different.

We are also interested in the number of items forming a bundle.



We can see that respondents did not show a strong preference for bundles of a certain size. The percentages in the graph represent the amount of time a bundle was selected out of all the times it was shown, so they don't sum to 100%.
To see if this distribution is different from the original distribution of the bundles, we use a Chi-square test with 3 degrees of freedom. In the whole questionnaire, bundles with 1 and 4 features were 20% of the total each, while bundles with 2 and 3 were 30% of the total each.

The p-value obtained is 0.164 meaning that the difference with the questionnaire distribution is meaningful, although the difference is not so big.
We are now interested in knowing what was inside of the selected bundles.

	Observed	Expected	$(O-E)^2/E$
1 el	38	30,8	1,364211
2 el	48	46,2	0,0675
3 el	35	46,2	3,584
4 el	33	30,8	0,146667

Q=5,162377

df=3 p-value=0,164

All the features appeared in the same frequency in the questionnaire. They appear in chosen bundles with the following frequencies:

- Additional purchase insurance: 6,5% (100 days extra) and 6.3% (200 days extra)
- Additional travel insurance: 16,7%
- Travel inconvenience insurance: 17,9%
- Credit protection insurance: 21,65%
- SMS alert: 17,06%
- Statement fee: 16,17%

Each feature has the same probability of being part of a bundle. If the bundles were selected randomly, we would expect to see the same percentages, while it's clear there is a pattern.
We can notice how the three "big ticket" items are among the most selected.
These are the most expensive bundle items: Travel Inconvenience Insurance, Credit protection insurance and Statement fee.

We counted how many big ticket items are part of selected bundles. The results are:

- No big ticket item: 22,44
- One big ticket item: 41,67
- Two big ticket item: 32,29
- Three big ticket item: 3,21
- Four big ticket item: 0

Most bundles have just one or two elements. We will use this information to create the optimal configuration for a bundle.

We are also interested in studying the interaction between the "Purchase insurance" attribute of the core product and the "Additional purchase insurance" present in the bundles.
In the questionnaire, bundles with additional purchase insurance were not proposed with cards that had no purchase insurance.
The results were the following:

Table 7.8: Link exra insurance-insurance in core product

extra purchase insurace in selected bundle	not present	100 days	200 days	1 year	total
not present	44	31	37	45	113
100 extra	0	11	11	11	33
200 extra	0	12	6	10	28
total		54	54	66	174

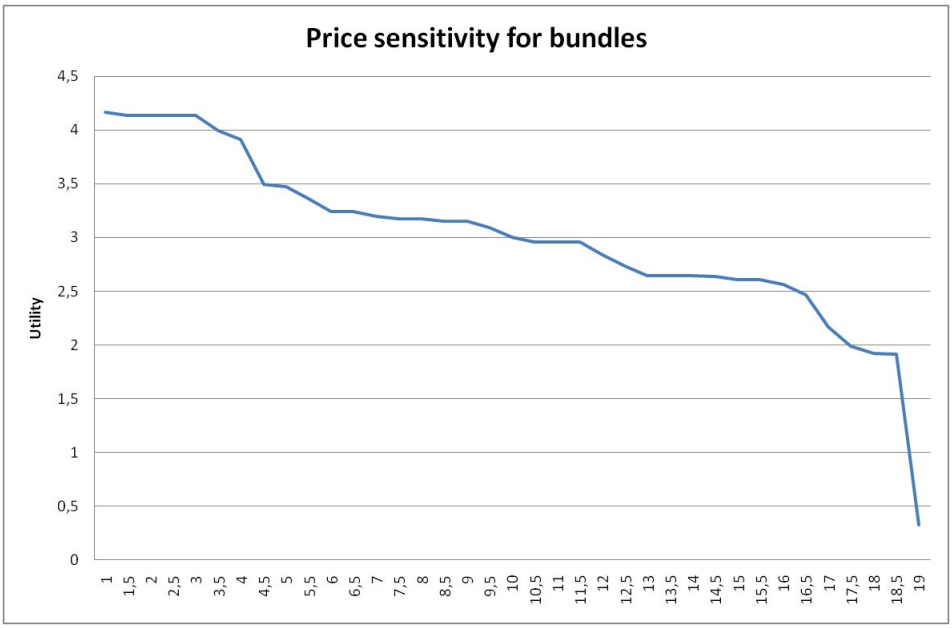
We were therefore interested in studying the link between the two. We tested independence between the two factors with a Chi-square test. The p-value is 0.55 meaning that the two factors are not independent.

	100 days	200 days	1 year
not present	35,06897	35,06897	42,86207
100 extra	10,24138	10,24138	12,51724
200 extra	8,689655	8,689655	10,62069

We also studied the price sensitivity for bundles. The biggest decreases is between 4 and 5. Another smaller decrease happens before 10. We can regard 5 and 10 as psychological price barriers.

	100 days	200 days	1 year
not present	0,472112	0,10633	0,106639
100 extra	0,056194	0,056194	0,183908
200 extra	1,261084	0,832512	0,036274

Q=3,1112473236367
p-value=0,211



Optimal configuration

With the insight gained with our analysis, we cam calculate the optimal configuration for a bundle offer.
This configuration is based on utilities and the analysis we carried on earlier.
First we have to decide how many bundles to offer. In the questionnaire we studied only one bundle per product, but we can use utilities to see what would happen when more bundles are offered.
It is possible to simulate bundle choices in the sheet *utils* of *bundle results.xlsm*.

After trying various combinations, our proposal is the following: two bundles, a cheap one and a middle-priced one, with no elements in common.
The best structure for this configuration is:

Bundle 1: travel protection insurance, credit protection insurance, SMS alert for 9.50 Euro
Bundle 2: statement fee for 4 Euro.

With this configuration, a total of 32 respondents (out of 49) would select at least one bundle.
Twenty-three would select the first bundle, twenty one would select the second. This mean that twelve respondents would select both of them.
From our point of view, since one bundle is quite cheap compared to the other, we can expect many people to select both of them. These two bundles contain all items a part from Additional Purchase Insurance, the least selected item in the study.

Chapter 8

CBC HB estimation algorithm in R

The algorithm presented in this chapter replicates the one implemented by Sawtooth Software CBC-HB.
It was decided to implement this algorithm in the open source language R so that everybody could replicate the results shown in this thesis without the need to pay for proprietary software.
This algorithm has its own way to code the input files but we wrote a VB code to pass from Sawtooth input format to R input format.
The code is partly based on the R package *bayesm*, written by Rossi E., Allenby G., McCulloch R., and freely available to download and modify under the GPL licence.
The package for *bayesm* is available at the site `http://cran.r-project.org/web/packages/bayesm/index.html`.

Our implementation of the CBC HB algorithm with a test example is contained in the archive attached to the thesis.

8.1 Running the program

To illustrate how to use the program, we consider a test run.

In this example 300 respondents partecipated in a survey.
Each respondent answered 12 choice tasks, each one with 5 alternatives.
Products were defined by four attributes, each one having seven levels, numbered 1 to 7.

8.1.1 Design matrix and answers

The design matrix is recorded in the object *lgtdata*, a list with a number of entries equal to the respondents number.
One entry, the one for respondent number 300, looks like this:

```
>lgtdata[[300]]
$y
 [1] 4 2 4 1 1 3 1 4 4 2 3 3

$X
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
 [1,]    0    0    0    0    0    1    0    0    0     0     0     1     1
 [2,]    0    0    0    0    0    1    0    0    0     0     0     1     0
 [3,]    0    0    0    0    0    1    0    0    0     0     0     1     0
 [4,]    0    0    0    0    0    1    0    0    0     0     0     1     0
 [5,]    0    0    0    0    0    0    0    0    0     0     0     0     0

....

[55,]    0    0    1    0    0    0    0    0    0     0     0     1     0
[56,]    0    0    0    0    0    1    0    0    1     0     0     0     0
[57,]    0    0    0    0    0    1    0    1    0     0     0     0     1
[58,]    0    0    0    0    0    1    1    0    0     0     0     0     0
[59,]    0    0    0    0    0    1    0    0    0     0     1     0     0
[60,]    0    0    0    0    0    0    0    0    0     0     0     1     0
      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
 [1,]     0     0     0     0     0     0     0     0     0     0     1
 [2,]     0     1     0     0     0     0     0     0     0     0     1
 [3,]     0     0     1     0     0     0     0     0     0     0     1
 [4,]     0     0     0     1     0     0     0     0     0     0     1
 [5,]     0     0     0     0     1     1     0     0     0     0     0

....

[55,]     0     1     0     0     0     1     0     0     0     0     0
[56,]     0     0     0     1     0     0     0     0     0     0     1
[57,]     0     0     0     0     0     0     0     0     0     0     1
[58,]     0     0     0     0     0     0     0     0     0     0     1
[59,]     0     1     0     0     0     0     0     0     0     0     1
[60,]     0     0     0     0     1     0     0     1     0     0     0

$hhid
[1] 300
```

The vector y contains the answers to the 12 questions.
The matrix X size is 24*60. Each horizontal line represents a product shown in

the choice task, with a 1 when that attribute is present in the product. Since there are 12 questions, each with 5 products, the total is 60 lines.
The number of columns is 24 since for each attribute one of the levels's utility is fixed as 0 in the computations.This also means that it is not necessary to write if that level is present in a certain product, since it doesn't add utility.

8.1.2 Prior parameters

This code runs the algorithm for 10.000 iterations, keeping values in one every 5 iterations to calculate parameter estimates.

The prior for this model are the standard uninformative priors, the default priors of the model:

```
#the standard (default) uninformative priors for this model:
#b_i ~ N(mu,Sigma).
#mu ~ N(mubar,Sigma*Amu^{-1})
#Sigma ~ IW(nu,V)

#mubar: prior mean vector (default: 0)
#Amu: prior precision for normal comp mean (def: 0.01 I)
#nu: deg. of freedom parameter (defauilt: nvar+3)
#V: pds location parm for IW (default: nul)
```

8.1.3 Running the algorithm

```
library(bayesm)

setwd('C:\\cbc opt test -het1E-')

#loads the functions used for the estimation
source('createXm.r')
#loads the questionnaire matrix X and the answers
#-----------------------
#vector containing the number of attributes per level
maxlevs=c(7,7,7,7)
#nresp number of respondents
nresp=300
#number of questions per respondent
numquest=12
#number of alt per question
numalt=5
#-----------------------
source('formatXm.r')
#formats the matrix X to be used for the computation
```

```
#-----------------------
#parameters for the estimation

#keep a draw every "keep" iterations
keep=5
#number of iterations
R=10000
mcmc1=list(keep=keep,R=R)
#-----------------------
#the standard (default) uninformative priors for this model:
#b_i ~ N(mu,Sigma).
#mu ~ N(mubar,Amu^{-1})
#Sigma ~ IW(nu,V)

#deltabar: vector of prior means (default: 0)
#mubar: prior mean vector
#Amu: prior precision for normal comp mean (def: 0.01 I)
#nu: deg. of freedom parameter (defauilt: nvar+3)
#V: pds location parm for IW (default: nul)
#ncomp: number of normal components for the mixture of normals

#run the estimation
source('functionsetM.r')
out=rhierMnlRwMixtureMod(Data=list(p=numalt,lgtdata=lgtdata,Z=NULL),
    Prior=list(ncomp=1),Mcmc=mcmc1)
#-----------------------
source('countexactm.r')
#calculates the prediction ratio (in variable 'ratio')
source('writeresults.r')
#writes the estimated betas in .csv format (by default in position c:\\estbeats.csv)
```

8.2 Output files

The results of the estimation are written in the *out* object.
The field $\$betadraw$ contains betas. The strucute is $out\$betadraw[respnum, levelnumber, keptiterationnumber$
The kept iteration number goes from 1 to $keep/R$.
So, in our example, if we want the betas of respondent 1 at the last iteration we would write $out\$betadraw[1, , 2000]$.

In the field $\$loglike$ contains the log likelihood of the model, calculated one every *keep* iterations.

Bibliography

[1] Bertolini M., Ofek E., Ariely D., *The effect of Add-On features on perceived product value*, London Business School working papers, 2008.

[2] Chernev A., Hamilton R., *Assortment Size and Option Attractiveness in Consumer Choice Among Retailers*, Journal of Marketing Research, volume XLVI.

[3] Dellaert B., Stremersch S., *Marketing Mass Customized Products: Striking the Balance between Utility and Complexity*, Journal of Marketing Research, August 2005.

[4] Evgeniou T., Pontil M., Toubia O., *A fast optimization approach to robust heterogeneity modeling for choice-based conjoint estimation*, `http://w4.stern.nyu.edu/ceder/research.cfm?doc_id=5402`, 2005.

[5] Hastie T., Tibshirami R., Friedman J. *The elements of statistical learning* . Springer, second edition, 2009.

[6] Iyengar R., Jedidi K., Kohli R., *A conjoint approach to multi-part pricing*, Journal of Marketing Research, Vol. XLV, No. 2, April 2008

[7] Koop G., *Bayesian Econometrics*. John Wiley & Sons, 2003.

[8] Liechty J, Ramaswamy V., Cohen S., *Bayesian econometrics*, John Wiley & Sons, 2003.

[9] Liechty J, Ramaswamy V., Cohen S., *Choice Menus for Mass Customization: An Experimental Approach for Analyzing Customer Demand with an Application to a Web-Based Information Service*, Journal of Marketing Research, Vol. XXXVIII (May 2001).

[10] Marder E., *The assumptions of choice modelling: Conjoint analysis and SUMM*, Canadian Journal of Marketing Research, Volume 18, 1999.

[11] Maxwell S. *The price is wrong*, John Wiley & Sons, 2007.

[12] Rossi E., Allenby G., McCulloch R., *Bayesian statistics and marketing*, John Wiley & Sons, 2005.

[13] Sawtooth Software, *The CBC/HB system for Hierarchical Bayes estimation - version 4.0*, http://www.sawtoothsoftware.com/, 2005.

[14] Solomon M., Bamossy G., Askegaard S., Hogg M., *Consumer behavior - an European perspective*, Prentice Hall, third edition, 2006.

[15] Train K., *Discrete Choice Models with Simulation*, Cambridge University Press, 2003.

[16] Wooldridge J.M., *Econometric analysis of cross section and panel data*, MIT Press, 2002.