Bachelor's thesis Econometrics & Operations Research

Comparing three different data collection methods in conjoint analysis

Author:
Marchella Maria
305097

Supervisor:
Dr. A.J. Koning

Co-reader:
Msc. K.Y. Lam

Summary

In conjoint analysis is selecting a data collection methods the one of the most important steps for a successfully analysis, therefore it is important to find the best data collection method. In this research we have examined the three data collection methods, (1) Best–scaling, where the respondent chooses the best product; (2) One pair best-worst scaling, where the respondent chooses the best and worst product and (3) Two pair best-worst scaling, where the respondent chooses 2 best and the 2 worst products. The collection methods will be analyzed were mobile phone is the product, in a conjoint analysis with partial ranking. Partial ranking is a new ranking method that eases the respondent burning for ranking. In this research we could not find a collection method that is significant better than the other. Because this problem is never been examined it gives information for a follow-up study.

Index

Chapter 1

Introduction

There are different ways to measure consumer judgements in the marketing sector. The last few decades' marketers found out that consumer's purchasing decision is based on several underlying attributes/features of a product. This results in a purchase or consumers willing to pay a higher price. For these reasons, it is important for producer's to know which product attributes are preferred by the target group. The main problem is that when consumers are asked which of the attributes they classify as important, consumers indicate that all attributes are important. Conjoint analysis is a quantitative research method that addresses this problem, whereby respondents are asked to make trade-off judgements.

Conjoint analysis, a research method developed in the 1970's see Green and Wild (1975) has 4 objectives; (1) to help select features to offer on a new or revised product or service; (2) to help set prices; (3) to predict the resulting levels of sales or usages; (4) try out a new-product concept. Conjoint analysis provides a quantitative measure of the relative importance of one attribute as opposed to another. Understanding precisely how people make decisions, producers can work out the optimum level of features and services that balance value to the customer against cost to the company.

Different researchers have successfully used conjoint analysis, while the research method is still renewing. Green and Srinivasan (1990) have written the paper 'Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice'. This article provides a basic overview of how conjoint measurement works and the prospects, this paper reviews the work of Green and adjustments made.

Different data collection methods are used in conjoint analysis. Louviere et al. (2008) identify and compare four data collection methods including (1) constant sum scaling of product attribute importance; (2) best-worst scaling of product attribute importance; (3) product attribute importance inferred from statistical effects in discrete choice experiments and (4) implied willingness to pay for difference in attribute levels. Another data collection method is rating scales, whereby respondents need to rate attributes. Of the different data collection methods, researchers often use Best-Worst Scaling. Best-Worst Scaling, also known as Maximum Difference Scaling, is descendent of the stated choice methods a kind of partial ranking. Where traditionally only best choice methods have been of interest to researches, Marley and Louviere (2005) have proven that Best-Worst Scaling, compared to best choice scaling, increases information efficiency.

Best-worst scaling has a number of advantages over traditional discrete choice tasks like rating scaling. (1) a single pair of best–worst choices contains a great deal of information about the person's ranking of options; (2) best–worst tasks take advantage of a person's propensity to identify and respond more consistently to extreme options; and (3) best–worst tasks seems to be easy for people. Despite the increasing use of the approach, the underlying models have not been examined, leaving practitioners without clear guidelines on appropriate experimental designs, data analyses, and interpretation of results (Marley and Louviere, 2005).

Previous literature has proved the disadvantages of rating scales. Researchers addressing the disadvantages of ratings scales as a data collection method are Crask and Fox (1987), Ben-Akiva et al. (1991), Bleichrodt and Johannesson (1997) and Goodman et al. (2006). All articles address two important occurring problems of rating scales: (1) their measurement properties; (2) their degree of discriminatory power. Measurement properties: scientists often assume that rating scales are interval scales with absolute difference between scale points; this assumption has been found to be frequently violated for ratings scales. Furthermore the degree of discriminatory power; rating data can have poor discriminatory powers, because each item is rated alone. Respondents must guess at the level to assign for the first item without having any items to compare.

For this reason, I will not compare rating scales versus Best-worst scaling in this paper. In this paper, I identify three data collection method variants of attribute stated choice methods.

1) Best –scaling, where the respondent chooses the best product
2) One pair best-worst scaling, where the respondent chooses the best and worst product
3) Two pair best-worst scaling, where the respondent chooses 2 best and the 2 worst products.

So the ***Research question*** *is:*

Which of the three data collection methods in a conjoint analysis

estimates better with a partial ranking model?.

Several researches address Best-Worst Scaling as a data collection method. Finne and Louviere (1992) propose a discrete choice task in which a person is asked to select both the best and the worst option in an available set of options. Since the publication of that paper, interest in, and the use of best-worst choice tasks as been increasing. Recent research by Marley and Louviere (2005) has proven that analysing the best and worst choices out of multiple items increase the information efficiency compared to only analysing the best choice. In the article of Mueller et al. (2009) a different variant of the Best-worst scaling that is part of the family of the stated choice methods is used. Participants were asked to classify, of eight different types of wine, the two best (most liked) and the two worst (least liked).

In conjoint analysis best worst ranking is a common data collection method and is estimated like any other data collection method with a multinomial logit model (MNL). Modeling a partial ranking model should not be estimated with a MNL model, because of the complexity of exact probabilities. The last two decades, modeling and estimation of ranking models are renewed; in the paper of K.Y.Lam et al.

(2010) they reduced the complexity of probabilistic ranking models by using approximate probabilities rather than exact ranking probabilities and concentrated on the Thurstone order statics models. K.Y.Lam et al. (2010) has shown that their models (Thurstone-Mosteller-Daniels model and a Luce model) are able to handle any kind of partial ranking. In this paper, the three different collecting methods will be estimated with the models used in K.Y.Lam et al. (2010).

Further in this paper; in chapter two the methodology of this paper will be outlined, the model will be explained and applied. Chapter three presents the results. And the last chapter four attends to the conclusion, discussion, limitations and recommendations.

Chapter 2

Methodology

2.1 Data

In conjoint analysis study a distinction is made between two phases: (1) data collection and (2) data analysis. This section reviews the data collection phase.

The framework of Vriens (1995) will be used to explain this phase.
Steps in the data collection phase

1) *Selection and definition of the attributes*
2) *Determination of relevant levels within attributes*
3) *Thinking about the preference model*
4) *The choice of the data collection method*
5) *Construction of the stimuli*
6) *Defining the dependent variable*
7) *The choice of the data collection procedure*

2.1.1 Selection and definition of the attributes

An *attribute* is a characteristic of a product (e.g. color), made up of various *levels* (there must be at least two for each attribute) or degrees of that characteristic (e.g. red, yellow, blue). The selection and definition of the attributes can be described in three questions: 1) which attributes should be included? 2) How many attributes should be included? 3) How should these attributes be defined? Normally, management and customers determine the first two questions. However, it is important to choose the number of attributes wisely, because it may influence the possibilities available in the subsequent steps in the Vriens-Framework. The third question needs to me understandable for customers; as for engineers attributes can be simple, yet for consumers attributes may sometimes be difficult to interpret. Therefore, attributes should be defined in terms of consumer benefits.

2.1.2 Determination of relevant levels within attributes

Determination of relevant levels within attributes can be described in two questions as well: (1) how much variation should exists in each attribute? And (2) how many levels should be defined within each attribute? The first question is for some attributes simple, because the range of variation is straightforward. For example, objects could differ in having a variation in the feature or not. For other attributes this can be more difficult. In case of continues attributes Vriens and Wittink (1992) argue that the range of variation is determined by the minimum amount of that attribute that the management cannot imagine going below, and the maximum amount management cannot imagine exceeding. If the range of an attribute is determined, than the numbers of levels can be determined. The number of levels you choose to define can have a significant bearing on the results. The first concern is called the “Number-of-Levels Effect.” When attributes defined on more attributes tend to get more importance. However, increasing the number of intermediated levels can lead to “attribute-level effect”. To explain, it is usually better to have more data at each price point than to have thinner measurements at more price points. Measuring too many points along a quantitative function can result in troublesome reversals. If you cover the entire range of interest with fewer levels, you can interpolate between levels within the market simulator to get finer granularity if needed.

2.1.3 Preference model

Should interaction effect be allowed? Once attributes and associated levels have been determined, analysts typically use some form of orthogonal design to generate different combinations of attribute levels called "profiles" (e.g., Green 1974; Louviere 1988a). A profile is a single attribute level combination in a complete factorial combination of attribute levels (called a “treatment combination” in the statistical design literature). For example, a given level of one attribute may only appear if another attribute has a given value, this is called the main effect plan. If the form of orthogonal design is not used there can be interaction effected between the different levels of different attributes. This results in main effects models that can lead to unknown and potentially large bias in the utility parameters that are estimated. For this reason interaction effects are not included in the preference model.

2.1.4 The choice of the data collection method

There are several alternative methods available for the collection of the conjoint data. 4 alternatives are discussed.

1) Full profile method
   Under this method respondents are asked to rank order, rate and etc., a set of objects that differ from one another on two or more attributes and that are defined on all attributed which are included in the study. The disadvantage of this method is the big possibility of information a task overload.

2) Trade-off method
   Under this method respondents are asked to rank each combination of levels of two attributes, from most preferred to least preferred. A big disadvantage is to construct a matrix in case there are more than 2 attributes, because if we got a n attributes we need to construct a (n(n-1))/2 matrices.

3) Paired comparison method
   Under this method respondents are asked to choose between two objects at a time, and choose the one preferred. These objects can be defined on all attributes included in the study or they

can be defined on a subset of all attributes. The disadvantage of the method is that the paired comparisons increase fast as the number of objects increase.

4) Attributed Stated choice method
Under this method respondents are as well asked to choose between objects, that are defined on all attributes in the study. This method can be divided in three groups.
   1) Best –scaling, where the respondent chooses the best object
   1) One pair best-worst scaling, where the respondent chooses the best and worst objects
   2) Two pair best-worst scaling, where the respondent chooses 2 best and the 2 worst objects.

2.1.5 Construction of the stimuli.

In case of Stated Choice Methods, these are the attribute levels constitute the pieces of information assumed to impact the decision maker's response and being manipulated by the experiment. That is why stimuli is constructed. However, Orme (2002) argues, "Manipulation should be used sparingly, or not at all. Specifying unnecessary or excessive manipulation is one of the most common mistakes. The problem usually begins when either the analyst (or the analyst's client) notices that some product combinations displayed during the interview are not realistic, given what currently exists in the market. Sometimes a product is shown with all the best features at the lowest price; or two attribute levels that would not naturally occur in the real world are paired together. The inclination is simply to prohibit such combinations".

Too many manipulations, in the best case, can lead to imprecise utility estimation and, in the worst case, confounded effects and the complete inability to calculate stable utilities. It is better to prompt respondents that they will see combinations during the interview that are not yet available in the market or that seem unlikely.

The specific combinations of the stimuli are given in the plan matrix, usually obtained by an experimental design. Each row of the plan matrix corresponds to a stimulus, and gives the values of the attributes. The attributes are allowed to be factors.

	Attribute 1	Attribute 2	. . . .	Attribute *m*
Stimulus 1			. . . .	
Stimulus 2			. . . .	
⋮	⋮	⋮		⋮
Stimulus c			. . . .	

2.1.6 Defining the dependent variable

Respondents were asked to evaluate the conjoint stimuli. Several types of evolution can be used to describe the dependent variable.

1) Allocations: the respondent will be asked to place or allocate the conjoint stimuli into categories.

2) Choices: a set of choice sets of the different stimuli is constructed and respondents will be asked to select one of the choice sets.

3) Rank ordering: the respondent will be asked to rank order the stimuli, this will be measured on ordinal scale.

4) Rating scales: the respondent will be asked to evaluate each conjoint stimulus on a rating scale (most liked to most unliked).

2.1.7 The choice of the data collection procedure

In marketing research there are different types of marketing tools that we can use to collect data, like mailing, telephone, personal interview, and briefing. In our case, students of the Erasmus University Rotterdam were asked during their class to fill in a questionnaire.

2.2 Data collection illustration

This part illustrates application of the data collection phase for the data we are going to use in this research further. The product used in this research is the mobile phone, the reason that we choose the mobile phone, is that it is a popular product. Everyone knows something about it and has got a opinion, over the question how should the perfect mobile phone looks like. Because the most mobile phone now and day are already extensive with allot of features, we specify the product to the Smartphone's. Because now allot of features are already in that type of phone.

2.2.1 Selection and definition of the attributes

To define the attributes we need to answers the following questions:

Which attributes should be included? At first it seems to be an easy question, just look at the different features of a mobile phone, like the size, colour, hardware, Bluetooth, qwerty-keyboard etc and select them. Because the most features already 'common' so the respondent does not recognise the importance of an attribute.

How many attributes should be included? Because it is a small research we had a limitation on the amount of the attributes. Therefore we use three attributes for each product.

How should these attributes be defined?

Attribute 1: hardware (touch screen or QWERTY keyboard); attribute 2: application (different kind of software) and attribute 3: extra features (something that specifies it with other smart phone).

2.2.2 Determination of relevant levels within attributes

The different variations in the attribute and the amount of levels in this case are two options: 1) the choice whether or not the attribute is included in the product and 2) attribute with 3 levels, to many levels can give number-of-levels Effect and can give the respondents complication when they fill the questionnaire in.

In this research we split the products in two groups, we did it because we want to see if the respondent has got less levels is there a different in the choice.

1) First group, where respondents had 3 attributes with 2-2-3 levels

Attribute→	Hardware	Application	Extra
Level 1	Qwerty-Keyboard	Free	GPS
Level 2	Touch screen	Paid	Wi-Fi
Level 3			Dual Sim

2) Second group, where respondents also had 3 attributes but now with all 2 levels.

Attribute→	Hardware	Application	Extra
Level 1	Qwerty-Keyboard	Free	GPS
Level 2	Touch screen	Paid	Wi-Fi

2.2.3 Preference model

Because interaction effect can lead to large bias in the parameter estimation we do not included interaction effects in the preference model.

2.2.4 The choice of the data collection method

In this research we used the Attributed Stated choice data collection method. Under this method respondents are as well asked to choose between objects, which are defined on all attributes in the study. This method is divided in three groups.

1) Best –scaling, where the respondent chooses the best object
2) One pair best-worst scaling, where the respondent chooses the best and worst objects
3) Two pair best-worst scaling, where the respondent chooses 2 best and the 2 worst objects

2.2.5 Construction of the stimuli

The stimuli of group 1

	Hardware	*Application*	*Extra*
Stimulus 1	Qwerty-Keyboard	Paid	Dual sim
Stimulus 2	Touch Screen	Paid	GPS
Stimulus 3	Qwerty-Keyboard	Free	GPS
Stimulus 4	Touch Screen	Free	Dual sim
Stimulus 5	Qwerty-Keyboard	Paid	Wi-Fi
Stimulus 6	Touch Screen	Paid	Wi-Fi
Stimulus 7	Qwerty-Keyboard	Free	Dual sim
Stimulus 8	Touch Screen	Free	GPS
Stimulus 9	Touch Screen	Free	Wi-Fi
Stimulus 10	Qwerty-Keyboard	Free	Wi-Fi
*Stimulus 11**	Touch Screen	Paid	Dual sim
*Stimulus 12**	Qwerty-Keyboard	Paid	GPS

The stimuli of group 2

	Hardware	*Application*	*Extra*
Stimulus 1	Qwerty-Keyboard	Paid	GPS
Stimulus 2	Touch Screen	Free	Wi-Fi
Stimulus 3	Qwerty-Keyboard	Free	GPS
Stimulus 4	Touch Screen	Paid	Wi-Fi
Stimulus 5	Qwerty-Keyboard	Paid	GPS
Stimulus 6	Touch Screen	Free	Wi-Fi
Stimulus 7	Touch Screen	Paid	Wi-Fi
Stimulus 8	Qwerty-Keyboard	Free	GPS

* After taken the questionnaires we found a typo. Stimulus 11 and 12 were missing and stimulus 1 and 2 were double on the questionnaires. Therefore in this research group 1 has got 10 stimuli instead of 12.

2.2.6 Defining the dependent variable

In our case, the Attributed Stated choice method, we use a mix of the different evaluations types. Different respondents were presented with one of the three options below. Best–scaling, where the respondent chooses the best object. One pair best-worst scaling, where the respondent chooses the best and worst objects. Two pair best-worst scaling, where the respondent chooses 2 best and the 2 worst object rankings. The dependent variable is in our case the partial ranking of each respondent different. In the chapter Statistical methodology they will be explained over partial ranking.

2.2.7 The choice of the data collection procedure

In our case, students of the Erasmus University Rotterdam were asked during their class to fill in a questionnaire. There were 6 different questionnaires and were random divided.

1) Best-scaling- group 1
2) One pair best-worst scaling group 1
3) Two pair best-worst scaling group 1
4)Best-scaling- group 2
5) One pair best-worst scaling group 2
6) Two pair best-worst scaling group 2

The questionnaires of these are given in Appendix.

2.3 Statistical methods

The previous section discussed the data collection phase, where the different types of stimuli were discussed. These results are used to construct a model. This section in mainly cited from K.Y.Lam et al. (2010). the methodology of the model that we used is described in more details.

2.3.1 Ranking

In the section Data was concluded that using partial ranking will be used for the dependent variable instead of full ordering. The difference between full ordering and partial ranking is that partial ranking implies a partial ordering of the stimuli. To explain the difference of these ranking, we start to explain the full ordering ranking using the stimuli of group 1 from the illustration.

Consider a single respondent who lists all stimuli *1,2,…,10* in order of preference, with the most preferred stimulus listed first. For each stimuli c in {1,2,…10} we define the rank $\pi_{(c)}$ of c as the position of c within this ordering. For example, $\pi_{(3)}=7$ indicates that stimulus 3 is listed in the 7th place in order of preference. We shall refer to $\pi = (\pi_{(1)}, \pi_{(2)},\ldots, \pi_{(10)})$ as a full ranking.

The full ranking π allows us to reconstruct the ordering. So for each rank *r* there exist exactly one *j* such that $\pi_{(j)} = r$; denote this stimulus by $\pi^{-1}_{(r)}$ so that $\pi^{-1} = (\pi^{-1}_{(1)}, \pi^{-1}_{(2),\ldots,} \pi^{-1}_{(10)})$. For example, $\pi^{-1}_{(7)}=3$ denotes that stimulus 3 is listed in 7th place in order of preference.

We assume that for each respondent the probability p_π of actually obtaining π as full ranking depends on a 10-dimensional linear predictor vector $\eta= (\eta_1, \eta_2,\ldots, \eta_{10})'$, that is $p_{\pi=}$ p(π| η). A so-called ranking model specifies the exact nature of the dependence of p_π on η. Analyzing the rankings can be done with different types of Thurstone order statics models like, Luce model or Thurstone-Mosteller-Danels model.

The problem in the analysis of rankings is handling the ties and the handling of missings. By missing we mean that a none rank is assigned to a stimulus, it may occur due to requirements imposed by the research design. A tie means that the same rank is assigned to multiple stimuli, it can occur due to the respondent's inability to differentiate between two or more stimuli. It is a common error, because the respondent finds it hard to compare too many choice options. Asking a respondent to rank only a subset of stimuli like Best-Worst ranking can solve this.

Therefore ranking which contains ties or missing should be considered as partial ordering of the stimuli rather than a full ordering, we call that partial ranking. Observe that for each partial ranking ϖ there exists a set of $S\varpi$ of all full rankings which do not contradict the partial ordering implied by ϖ. The probability assign by $p_{\varpi} = p(\varpi|\eta) = \sum_{\pi \in S\varpi} p(\varpi|\eta)$ to the partial ranking ϖ.

2.3.2 Model

With a view to perform a statistical analysis, we can see the model as a linear model for conjoint experiments with partial rankings . This probability model with probability $p\varpi$ has got the same structure as weighted least regression.

In order to the analysis we now focus on the question how the attributes of the stimuli influence the predictor vector η. In the previews chapter we already established that every stimulus has a specific combination of levels of the attributes, therefore x_{cm} denote the value attribute m takes for stimulus c, and we assume that the stimuli are describes by M attributes. Also we assume that predictor vector η= $(\eta_1, \eta_2,\dots, \eta_c)'$ is given by $\eta_c = \beta_1 x_{c1} + \beta_2 x_{c2} + \beta_M x_{cM} = \sum_{m=1}^{M} \beta_m x_{cm}$

Where $\beta_1, \beta_2,\dots, \beta_m$ are unknown coefficients. We can also write η=**X**β , where β is the M dimensional coefficient vector $(\beta_1, \beta_2,\dots, \beta_m)'$ and **X** we call that the plan matrix, is the C x M matrix which contains the value x_{cm} in its (c,m) location.

To approximate the probability model it result that $p_{\overline{\varpi}} \approx \frac{\sum_{\pi \in S\varpi} \exp\{\boldsymbol{q}_\pi^t \boldsymbol{X}\beta\}}{\sum_{\pi'} \exp\{\boldsymbol{q}_\pi^t \boldsymbol{X}\beta\}}$

Where q_π is the C-dimensional vector containing the π_c^{th} expected score $q_{\pi(c):c}$ as c^{th} element. q_π is also named the rank matrix. For example, for $\pi_{(3)}=7$, the 3^{th} element of vector q_π is the 7^{th} excepted score $q_{7:c}$

Estimation of the probability model with partial ranking, can be done with maximum likelihood method. So we approximate $p_{\pi j}$ by $p_{\overline{\varpi}}$ and we estimate β by maximizing the corresponding approximate log-likelihood

$$\ln \tilde{L}(\beta) = \sum_{j=1}^{J} \left(\sum_{\pi \in S\varpi_j} \exp\{\boldsymbol{q}_\pi^t \boldsymbol{X}\beta\} \right) - J \ln \sum_{\pi'}^{J} \left(\sum_{\pi \in S\varpi_j} \exp\{\boldsymbol{q}_{\pi'}^t \boldsymbol{X}\beta\} \right)$$

Where the rankings $\varpi_1, \varpi_2,\dots, \varpi_J$ are independently obtained from J different respondents.

After the log likelihood is approximated, we used the Akaike's information criterion (**AIC**) to measure the goodness of fit of the estimated model. AIC is a tool for model selection, the model with the lowest AIC is the best model.

Definition $\boldsymbol{AIC} = 2k - 2\ln(L)$ where *k* is the number of parameters in the statistical model and *L* is the maximized value of the likelihood function for the estimated model.

2.3.3 Illustration of the model

This section illustrates application of the model, we are going to use the data of group 1 with best-worst scaling as data collection method.

The plan matrix **X** for group 1 is;

$$\mathbf{X}=\left\{\begin{matrix} & hardware & application & GPS & Wifi & dual\ sim \\ stimulus\ 1 & 0 & 1 & 0 & 0 & 1 \\ stimulus\ 2 & 1 & 1 & 1 & 0 & 0 \\ stimulus\ 3 & 0 & 0 & 1 & 0 & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ stimulus\ 10 & 0 & 0 & 0 & 1 & 0 \end{matrix}\right\}$$

Each row of this matrix corresponds to a stimulus and gives the values of the attributes. The attributes in this matrix are dummy variables instead of factors. When the attribute hardware takes the value 0 the feature will be qwerty-keyboard and when the attribute application takes the value 0 the feature will be free application. So, stimulus 3 has as features Qwerty-Keyboard, free application and GPS. The predictor for stimulus 3 is then $\eta_3 = \beta_1 x_{31} + \beta_2 x_{32} + \beta_3 x_{33}$

The rank matrix $\mathbf{q}_\pi$ for group 1, with best worst scaling as data collection method is;

$$\mathbf{q}_{\pi}=\left\{\begin{matrix} \boldsymbol{stimulus} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} & \mathbf{7} & \mathbf{8} & \mathbf{9} & \mathbf{10} \\ judge\ 1 & 1 & 3 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ judge\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 3 & 1 & 2 & 2 \\ judge\ 3 & 3 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ judge\ 48 & 2 & 2 & 1 & 2 & 2 & 2 & 2 & 2 & 3 & 2 \end{matrix}\right\}$$

Each row of this matrix corresponds to exactly one judge, and contains for each stimulus the corresponding partial ranks assigned by the judge. The smaller the partial rank, the more preferred the stimulus. In the $\mathbf{q}_\pi$ rank matrix judge 1 chooses stimulus 1 as best option and stimulus 2 as least preferred one.

Chapter 3

Results

In this chapter we will be discussing the results of the 6 different data sets. In the section data collection illustration, the data is described. First, we look at the difference between the three different data collecting methods in conjoint analysis with partial ranking of 10 mobile phones, of group 1. And then, we will look at the difference between the three different data collecting methods in conjoint analysis with partial ranking of 8 mobile phones of group 2.

A general remark for interpreting the results is that the estimated coefficients are inversely related to preference. Hence, the smaller its rank, the more preferred a stimulus is. That means that, a positives coefficient indicates that higher levels lead to a higher, i.e. worse, ranking. The size of preference can be described in the absolute value of the coefficient. Therefore, a positive coefficient value does not necessarily mean that the respective attribute is rejected, but that it is less preferable than the reference level.

3.1 Results of Conjoint Analyses of 10 mobile phones

3.1.1 Results of Best –scaling with 10 mobile phones

Attribute	Variable	Coefficient	Stand. Err.	p-value
Hardware	touch screen	-0.048	0.186	0. 800
Applications	paid	0.575	0.208	0. 006
Extra	GPS	-0.632	0.283	0. 025
	Wi-Fi	-0. 895	0.274	0. 001

n=50 Log likelihood value: -447.8649 AIC: 903.730

Table 1: Estimated coefficients

Table 1 shows the estimated coefficients. The reference levels are: qwerty-keyboard, free applications and dual sim. On the attribute hardware of the mobile phone, nothing can be said between the level difference of touch screen and qwerty-keyboard, because the coefficient level touch screen is not significant. Therefore there it makes no difference if a mobile phone has a touch screen or qwerty-keyboard. Paid applications on a mobile phone is the most preferred kind of application on a mobile phone according to the respondents, as free application is least preferred by the respondents. For the attribute extra feature, respondents preferred a mobile phone that can access the web with Wi-Fi the most, then they prefer than a mobile phone that has as extra feature GPS and at least the respondents want to buy a mobile phone with dual sim.

3.1.2 Results of Best-Worst –scaling with 10 mobile phones

Attribute	Variable	Coefficient	Stand. Err.	p-value
Hardware	touch screen	0.118	-0. 153	0. 439
Applications	Paid	1.035	0.178	0. 000
Extra	GPS	-0.014	0.195	0. 945
	Wi-Fi	-0. 722	0.191	0. 001

n= 48 *Log likelihood value: -371.4315* *AIC: 750.863*

Table 2: estimated coefficients

Table 2 shows the estimated coefficients. The reference levels are: qwerty-keyboard, free applications and dual sim. On the attribute hardware of the mobile phone, nothing can be said between the level difference of touch screen and qwerty-keyboard, because the coefficient level touch screen is not significant. Therefore there it makes no difference if a mobile phone has a touch screen or qwerty-keyboard. The coefficient of paid application is positive, therefore paid applications on a mobile phone is the least preferred kind of application on a mobile phone according to the respondents, as free application is the most preferred by the respondents. For the attribute extra feature is Wi-Fi's coefficients the most negative number. Therefore the respondents preferred a mobile phone that can access the web with Wi-Fi the most, more than a mobile phone that has as extra feature dual sim. Over the level GPS, nothing can be said because it is not significant. Therefore there it makes no significant difference between GPS and Wi-Fi or GPS and dual sim.

3.1.3 Results of 2Best-2Worst –scaling with 10 mobile phones

Attribute	Variable	Coefficient	Stand. Err.	p-value
Hardware	touch screen	0.526	0. 165	0. 001
Applications	paid	1.733	0.214	0. 000
Extra	GPS	-0.495	0.206	0. 016
	Wi-Fi	1. 636	0.231	0. 000

n= 46 *Log likelihood value: -447.8649* *AIC:903.730*

Table 3: estimated coefficients

Table 3 shows the estimated coefficients. The reference levels are: qwerty-keyboard, free applications and dual sim. Qwerty-Keyboard it the most preferred hardware type of a mobile phone according to the respondents as than touch screen the least preferred. The coefficient of paid application is positive, therefore paid applications on a mobile phone is the least preferred kind of application on a mobile phone according to the respondents, as free application is the most preferred by the respondents.

The coefficient of paid application is negative, therefore paid applications on a mobile phone is the most preferred kind of application on a mobile phone according to the respondents, as free application is least preferred by the respondents. For the attribute extra feature, respondents preferred a mobile phone with GPS the most, then they prefer a mobile phone with dual sim and at least the respondents prefer a mobile phone that can access the web with Wi-Fi.

3.2 Results of Conjoint Analyses of 8 mobile phones

3.2.1 Results of Best –scaling with 8 mobile phones

Attribute	Variable	Coefficient	Stand. Err.	p-value
Hardware	touch screen	0. 038	0. 230	0. 870
Applications	paid	1. 516	0. 332	0. 000
Extra	Wi-Fi	-0. 981	0. 268	0. 000

n=51 Log likelihood value: -156.8839 AIC: 319.768

Table 4: Estimated coefficients

Table 4 shows the estimated coefficients. The reference levels are: qwerty-keyboard, free applications and GPS. On the attribute hardware of the mobile phone, nothing can be said between the level difference of touch screen and qwerty-keyboard, because the coefficient level touch screen is not significant. Therefore there it makes no difference if a mobile phone has a touch screen or qwerty-keyboard. Free applications on a mobile phone is the most preferred kind of application on a mobile phone according to the respondents, as paid application is least preferred by the respondents. For the attribute extra feature, respondents preferred a mobile phone that can access the web with Wi-Fi, more than a mobile phone that has as extra feature GPS.

3.2.2 Results of Best-Worst –scaling with 8 mobile phones

Attribute	Variable	Coefficient	Stand. Err.	p-value
Hardware	touch screen	1.208	0. 208	0. 000
Application	paid	-1.446	0. 224	0. 000
Extra	Wi-Fi	1. 948	0. 262	0. 000
n= 53	*Log likelihood value: -347.3833*		*AIC: 700.767*	

Table 5: estimated coefficients

Table 5 shows the estimated coefficients. The reference levels are: qwerty-keyboard, free applications and GPS.Qwerty-Keyboard it the most preferred hardware type of a mobile phone according to the respondents as than touch screen the least preferred. The coefficient of paid application is negative, therefore paid applications on a mobile phone is the most preferred kind of application on a mobile phone according to the respondents, as free application is least preferred by the respondents. For the attribute extra feature, respondents preferred a mobile phone with GPS more, than a mobile phone that can access the web with Wi-Fi.

3.2.3 Results of 2Best-2Worst –scaling with 8 mobile phones

Attribute	Variable	Coefficient	Stand. Err.	p-value
Hardware	touch screen	-0.308	0. 160	0. 055
Application	Paid	1.305	0. 196	0. 000
Extra	Wi-Fi	0.758	0.171	0. 000
n=39	*Log likelihood value: -384.436*		*AIC: 774.872*	

Table 6: estimated coefficients

Table 6 shows the estimated coefficients. The reference levels are: qwerty-keyboard, free applications and GPS. If we look at a 5% significance level, than on the attribute hardware of the mobile phone, nothing can be said between the level difference of touch screen and qwerty-keyboard, because the coefficient level touch screen is not significant. Therefore there it makes no difference if a mobile phone has a touch screen or qwerty-keyboard. But if we look at a 10% significance level, than we see that respondents prefer touch screen hardware on their mobile phone more than a qwerty-keyboard. Free applications on a mobile phone is the most preferred kind of application on a mobile phone according to the respondents, as paid application is least preferred by the respondents. For the attribute extra feature, respondents preferred a mobile phone with GPS more, than a mobile phone that can access the web with Wi-Fi.

Chapter 4

Conclusion

In conjoint analysis a good collection methods is one of the biggest steps for a successfully analysis, therefore it is important to find the best collection method. In this research we have examined the three collection methods. Comparing the three different collection methods of 8 mobile phones we found allot of inconsistency. We discuss the various collection methods for each attribute.

- The first inconsistency is about the attribute hardware, for the model with the Best-scaling as collection method we did not found any significant evidence that respondents prefer qwerty-keyboard above a touch screen or vice versa. In the Best-Worst scaling collection method the respondents prefer a qwerty-keyboard above a mobile phone with touch screen. That is in contrast to, the model that used the 2Best-2Worst collection method where the respondents preferred a mobile phone with touch screen then a qwerty-keyboard.

- The second inconsistency is about the attribute application, where the respondents prefer free application more than paid application on their mobile phone with the Best-scaling collection method and the 2Best-2Worst-scaling collection method. Unlike the Best-Worst scaling collection method, where the respondents choose a mobile phone with paid application above a mobile phone with free application.

- The last inconsistency about the collection methods of 8 mobile phones, is the attribute extra feature. For the model with Best-scaling as collection method respondents prefer Wi-Fi more GPS as extra feature on their mobile phones. As opposed to, Best-Worst and 2Best-2Worst collection methods where respondents choose GPS as extra feature on the mobile phone more than Wi-Fi as extra feature.

These results show a remarkable outcome; if a respondent has to choose the mobile phone with the Best-scaling as collection method, than he favors a mobile phone with free application and Wi-Fi. It does not matter if it as touch screen or qwerty-keyboard as hardware. While if a respondent has to choose the mobile phone with the Best-Worst scaling as collection method, than he prefers a mobile phone with qwerty-keyboard, paid application and GPS the most. And when a respondent has to choose the mobile phone with the 2Best-2Worst scaling as collection method, than they prefer a mobile phone with touch screen, free application and GPS the most. This results show us that the respondents are inconsistent in choosing the mobile phone.

Comparing the three different collection methods of 10 mobile phones we also found some inconsistency but less than the group 1 with 8 mobile phones.

- First, we look at the attribute hardware for the models Best-scaling and Best-Worst scaling as collection methods there were no significant evidence that respondents prefer qwerty-keyboard above a touch screen or vice versa. In the 2Best-2Worst scaling collection method the respondents prefer a mobile phone with qwerty-keyboard more than a mobile phone with touch screen.

- Second inconsistency is about the attribute application, this is the only attribute that were the respondents of all three collection methods give the same outcome. So the respondents prefer free application more than paid application on their mobile phone.

- The third inconsistency, is about the attribute extra feature. For the model with Best-scaling as collection method respondents prefer Wi-Fi the most, then GPS and then dual sim. The respondents of the model with collection method Best-Worst scaling prefer Wi-Fi more than dual sim, GPS was in this model not significant. As opposed to the 2Best-2Worst collection method, where respondents choose GPS as the most preferred extra feature on the mobile phone, were dual sim and Wi-Fi the least favored extra feature.

The result that give the most inconsistent answers in group 2, is that in attribute extra feature. The level Wi-Fi, for the models Best-scaling and Best-Worst scaling as collection methods, the respondents choose that as most preferred feature on their mobile phone. In contrast to 2Best-2Worst scaling collection method where the respondents preferred Wi-Fi the least.

To answer the research question, *Which of the three data collection methods in a conjoint analysis estimates better with a partial ranking model?*

- First, because the coefficients are not stable for any model we could not compare the different model with each other. Unstable coefficients means that the coefficients were not close together, and they were not significant in any model. If the result where consistent you could compare the models with different test, e.g. likelihood ratio test.
- Second, the model with the lowest AIC is the best model. In the section results, the AIC of each model is given. The AIC as few conditions; (1) the different models should have the same number of data points and (2) the different models are descents from the same dataset. At both points the conditions fails, because in the result you can see that every model has a different *n*. And each model has its own dataset, because the rank matrix $\mathbf{q}_{\pi}$ is for each data collection method different.

Therefore we could not find statistical prove that one of the three collection methods is better than another. If I should choose a data collection method, I must choose the 2best 2worst scaling, because the judges are giving there more information than best-worst scaling and best-worst scaling gives more information than best-scaling as data collection method. If we argue in that way, we will go back to full ranking as data collection method, but we know from previous research that too much information can lead to noise. The perfect data collection method should be a trade-off between information and noise, but that is something for a follow-up study.

Bibliography

M. Ben-Akiva, T.Morikawa and F. Shiroishi. Analysis of the reliability of preference ranking data. *Journal of Business Research Volume 23, Issue 3, P 253-268, 1991*

H. Bleichrodt and M. Johannesson Standard gamble, time trade-off and rating scale: Experimental results on the ranking properties of QALYs. *Journal of Health Economics Volume 16, Issue 2, P 155-175, 1997*

Cohen, Steve and Leopoldo Neira .,2003. Measuring Preference for Product Benefits Across Countries: Overcoming Scale Usage Bias with Maximum Difference Scaling (ESOMAR 2003)

M.R. Crask and R.J. Fox. An exploration of the interval properties of three commonly used marketing research scales: a magnitude estimation approach. *Journal of the Market Research Society 29, P 317–339 ,1987*

S.Goodman, L. Lockshin and E. Cohen . Using the best–worst method to examine market segments and identify different influences of consumer choice. *In Proceedings of the third international wine business and marketing research conference*. *2006*

Finn, Adam and Jordan J. Louviere., 1992. "Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety," Journal of Public Policy and Marketing, Vol. 11, No. 2 pp.12-25.

P.E. Green, V. Srinivasan. Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice'. *Journal of Marketing 54*, *1990*.

P.E. Green and Y. Wind. New Way to Measure Consumers' Judgments. *Harvard Business Review* 53, p107-117, 1975.

K.Y. Lam, A.J. Koning and P.H Franses.2010. Ranking models in conjoint analysis. *Econometric Institute Report EI 2010-51*

J.J. Louviere and T. Islam. A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling. *Journal of Business Research 61 p903–911, 2008*

Louviere, Jordan., 1991. Best-Worst Scaling: A Model for the Largest Difference Judgements, Working Paper, University of Alberta, Canada.

A.A.J. Marley and J.J. Louviere. Some probabilistic models of best, worst, and best–worst choices. *Journal of Mathematical Psychology 49, p 464–480, 2005.*

M.Vriens*. Conjoint analysis in marketing: developments in stimulus representation and segmentation methods. ISBN: 9789072591265, 1995*

M. Vriens and D. Wittink *(1992), Data Collection in Conjoint Analysis, unpublished manuscript.*

Appendix

Questionnaires

Age: **gender**: male/female

Indicate which of the following mobile phones you prefer to use.

The mobile phone that I preferred the most…………………………………………..

Mobile phone	Hardware	Application	Extra
1	Qwerty- keyboard	Paid	Dual sim
2	Touch screen	Paid	GPS
3	Qwerty- keyboard	Open source	GPS
4	Touch screen	Open source	Dual sim
5	Qwerty- keyboard	Paid	Wi-Fi
6	Touch screen	Paid	GPS
7	Touch screen	Paid	Wi-Fi
8	Qwerty- keyboard	Open source	Dual sim
9	Touch screen	Open source	GPS
10	Qwerty- keyboard	Paid	Dual sim
11	Touch screen	Open source	Wi-Fi
12	Qwerty- keyboard	Open source	Wi-Fi

Age: **gender**: male/female

Indicate which of the following mobile phones you prefer the most and the least to use.

The mobile phone that I preferred the **most**…………………………………………..

The mobile phone that I preferred the **least**…………………………………………..

Mobile phone	**Hardware**	**Application**	**Extra**
1	Qwerty- keyboard	Paid	Dual sim
2	Touch screen	Paid	GPS
3	Qwerty- keyboard	Open source	GPS
4	Touch screen	Open source	Dual sim
5	Qwerty- keyboard	Paid	Wi-Fi
6	Touch screen	Paid	GPS
7	Touch screen	Paid	Wi-Fi
8	Qwerty- keyboard	Open source	Dual sim
9	Touch screen	Open source	GPS
10	Qwerty- keyboard	Paid	Dual sim
11	Touch screen	Open source	Wi-Fi
12	Qwerty- keyboard	Open source	Wi-Fi

Age: **gender**: male/female

Indicate 2 mobile phone which you prefer the most and the least to use.

The mobile phones that I preferred the **most**……………………**and**…………………..

The mobile phones that I preferred the **least**……………………**and**…………………..

Mobile phone	**Hardware**	**Application**	**Extra**
1	Qwerty- keyboard	Paid	Dual sim
2	Touch screen	Paid	GPS
3	Qwerty- keyboard	Open source	GPS
4	Touch screen	Open source	Dual sim
5	Qwerty- keyboard	Paid	Wi-Fi
6	Touch screen	Paid	GPS
7	Touch screen	Paid	Wi-Fi
8	Qwerty- keyboard	Open source	Dual sim
9	Touch screen	Open source	GPS
10	Qwerty- keyboard	Paid	Dual sim
11	Touch screen	Open source	Wi-Fi
12	Qwerty- keyboard	Open source	Wi-Fi

Age: **gender**: male/female

Indicate which of the following mobile phones you prefer to use.

The mobile phone that I preferred the most…………………………………………..

Mobile phone	**Hardware**	**Application**	**Extra**
A	Qwerty- keyboard	Paid	GPS
B	Touch screen	Open source	Wi-Fi
C	Qwerty- keyboard	Open source	GPS
D	Touch screen	Paid	Wi-Fi
E	Qwerty- keyboard	Paid	Wi-Fi
F	Touch screen	Open source	GPS
G	Touch screen	Paid	GPS
H	Qwerty- keyboard	Open source	Wi-Fi

Age: **gender**: male/female

Indicate which of the following mobile phones you prefer the most and the least to use.

The mobile phone that I preferred the **most**…………………………………………..

The mobile phone that I preferred the **least**…………………………………………..

Mobile phone	**Hardware**	**Application**	**Extra**
A	Qwerty- keyboard	Paid	GPS
B	Touch screen	Open source	Wi-Fi
C	Qwerty- keyboard	Open source	GPS
D	Touch screen	Paid	Wi-Fi
E	Qwerty- keyboard	Paid	Wi-Fi
F	Touch screen	Open source	GPS
G	Touch screen	Paid	GPS
H	Qwerty- keyboard	Open source	Wi-Fi

Age: **gender**: male/female

Indicate 2 mobile phone which you prefer the most and the least to use.

The mobile phones that I preferred the **most**……………………**and**…………………..

The mobile phones that I preferred the **least**……………………**and**…………………..

Mobile phone	**Hardware**	**Application**	**Extra**
A	Qwerty- keyboard	Paid	GPS
B	Touch screen	Open source	Wi-Fi
C	Qwerty- keyboard	Open source	GPS
D	Touch screen	Paid	Wi-Fi
E	Qwerty- keyboard	Paid	Wi-Fi
F	Touch screen	Open source	GPS
G	Touch screen	Paid	GPS
H	Qwerty- keyboard	Open source	Wi-Fi