

An Empirical Comparison of Willingness to Pay Methods within a Hypothetical Purchase Context

Ioannis Petrakis

Study Program: Management Economics and Consumer studies

Specialization: Marketing and Consumer Behaviour

Course Code: MCB-80436

Registration Code: 851206-650-060

Wageningen University

Marketing and Consumer
Behaviour Group

Supervisor: Dr Ivo van der
Lans

Contents

0. Abstract.....	3
1. Introduction	4
1.1 Preference data methods and WTP	4
2.2 Motivation of research	6
1.3 Research Questions.....	7
1.4 Outline.....	7
2. The process of determining WTP from preference data	7
2.1 Define the problem	8
2.2 Select the method.....	8
2.3 Design the survey	9
2.4 Collect the data	9
2.5 Determine WTP	9
3. Description of the hypothetical WTP methods in market research	10
3.1 Conjoint Analysis	10
3.2 Paired Comparison-Based Preference Measurement	13
3.3 Contingent Valuation	19
4. Scientific debate for the hypothetical WTP methods	20
4.1 Conjoint Analysis	20
4.2 Contingent Valuation	20
4.3 Paired Comparison Based Preference Measurement	21
5. Materials and Methods	21
5.1 Stimulus	21
5.2 Study Design	24
5.2.1 Choice Based Conjoint Analysis:	24
5.2.2 Paired Comparison-Based Preference Methodology:	25
5.2.3 Contingent Valuation:	25
5.3 Data Collection	25
5.3 Determination of mean WTP and creation of demand curves.....	26
5.3.1 Choice Based Conjoint Analysis	26

5.3.2 Paired Comparison-Based Preference Measurement	29
5.3.3 Contingent Valuation:	29
6. Results	29
6.1 Mobile phone.....	29
6.2 PC mouse	34
7. Discussion.....	38
8. Conclusion	40
Appendix	41
Appendix A.....	41
Appendix B	42
Appendix C	43
Appendix D.....	44
References.....	44

0. Abstract:

This study compared empirically the willingness to pay (WTP) determined of three hypothetical methods: The Choice Based Conjoint Analysis (CBCA); Contingent Valuation (CV); and Paired Comparison Based Preference Measurement (PCPM). Using a three-in-one design the relative performance of the methods on determining mean WTP and predicting demand was examined. The WTP was determined for a PC mouse and a mobile phone. For the mobile phone the results indicated that CBCA exhibited the most accurate WTP distribution and realistic mean WTP. For the PC mouse the results indicated that CV exhibited the most accurate WTP distribution and PCPM the most realistic mean. Also, discussion is taking place in the rest of the paper regarding the way that technical characteristics of the methods affected their performance on measuring WTP. Then the limitations of this study and recommendations for further research are provided.

Keywords: WTP, CBCA, CV, PCPM, predicted demand

1. Introduction

1.1 Preference data methods and WTP

The elicitation of preferences is very crucial for understanding and predicting consumer behaviour. One of the most common applications concerning the elicitation of preferences is the determination of willingness to pay (WTP) for a product (Miller et al., 2011; Grunert et al., 2009; Wertenbroch and Skiera, 2002; Alberini et al., 2004; Welsh and Poe, 1998). WTP has been defined by Wertenbroch and Skiera (2002) as “the maximum price a buyer is willing to pay for a given quantity of a good”, and it is a matter of great importance for marketing practices (Grunert et al., 2009).

According to Grunert et al. (2009) WTP information is vital for checking the expected profitability for new product development, product improvement, product differentiation, and line extensions. Also, Shafter and Zhang (1995) signify that WTP is critical for promotional activities. Moreover, Anderson et al. (1993) and Miller et al. (2011) denote that the correct measurement of WTP is crucial for shaping competitive strategies and implementing value audits. Further, knowledge of WTP is instrumental for the estimation of demand and the formation of optimal pricing schedules (Wertenbroch and Skiera, 2002). Consequently, there is a lot of active research on the development and testing of the various methods that are being used to determine WTP through the elicitation of consumers' preferences (Cameron and James 1987; Jedidi and Zhang 2002).

Along with the intense research, there is a lot of scientific debate regarding the trustworthiness of these methods. Concerning to the majority of the papers in the extant literature, the mean scores and demand curves of the alternative WTP methods differ significantly (Grunert et al., 2009; Miller et al., 2011; Wertenbroch and Skiera, 2002 among many others). Similarly to Wertenbroch and Skiera (2002), the term demand curve refers to a cumulative distribution graph which is created by individual WTPs derived from a single method and represents the relationship between the price of a certain product, say p , and the proportion of respondents who are willing to pay an amount of money bigger than p ($WTP > p$) in order to acquire that product. Demand curves are very important in WTP studies since they provide useful information regarding the observed demand, for the underlying product, across the sample of respondents who are examined under a specific method.

Two major distinctions among WTP methods are whether they measure WTP directly or indirectly and whether they determine consumers' hypothetical WTP or actual WTP (Miller et al., 2011). Table 1 provides a detailed classification of the alternative methods which are used for the determination of WTP. Since WTP is a context-dependent construct (Thaler, 1985), WTP methods can be primarily categorized according to how well they approximate the actual purchasing context of the product in question. The one is called actual WTP methods and the other hypothetical WTP methods (Louviere et al., 2000). Actual WTP methods are an expansion of hypothetical ones. In the latter the respondents only state their preferences about a priced product, whereas in the former after stating their preferences they are asked to reveal them by purchasing that product.

What majorly distinguishes actual and hypothetical WTP methods is the quality of the data. Hypothetical WTP data are gathered solely via simple questionnaires by asking the respondents to state their preferences for a hypothetical product. Actual WTP data are gathered through sophisticated research designs like simulated market transactions (Silk and Urban 1978), experiments (Vickrey, 1961; Becker et al., 1964) and incentive-align surveys (Miller et al., 2011; Ding, 2007). Actual WTP data techniques, which are the cornerstone of pricing research, have the assumed advantage of high external validity since consumers reveal their WTP by making an actual purchase. However, despite their realism, even these methods cannot elicit the maximum price the consumer is willing to pay.

According to Wertenbroch and Skiera (2002), the true WTP still remains unknown in actual WTP data methods. The data that is being derived through an actual purchasing context is able to uncover a price which applies to real world. But, even in real case scenarios the true WTP still remains unknown since it may not reflect the maximum price that

consumers have in mind. Actual data methods demand for the respondents to buy the product in question if their stated WTP is bigger than a specific threshold. So, respondents tend to state a lower WTP than their real one in order to avoid buying the product at the end of the study.

Therefore, no method that is included in the diagram of Figure1 is fool proof. The premise of this argument is that WTP can vary by individual and by different buying situation which makes its elicitation a very ambiguous task. In other words, even the most advanced techniques cannot accurately predict the natural WTP. Additionally, actual WTP methods cannot provide any support for products that they do not exist in the market yet (Grunert et al, 2009). According to Louviere et al. (2000) actual WTP data are time consuming and expensive to collect given the competitive pressure and complexity that exist in the current turbulent markets. Hence, one may opt for alternative methods which are easier and less costly to conduct.

On the other hand, hypothetical WTP methods, which are based solely on survey data, they can elicit consumer preferences when market data are not available or the product is not finalized. Actually, the most intrinsic advantage of hypothetical WTP methods is that they can catch up with the rapid shifts of technological frontiers because they are not limited to the current market and technology formation (Louviere et al., 2000). However, hypothetical WTP methods are often subject to criticism due to hypothetical bias (Hoffman et al., 1993) which means that respondents tend to give higher WTPs, in comparison with those that they would give in a real purchase context, since they do not have any budget constraint. In other words, economic commitment is absent in hypothetical methods which makes the suspicious for low external validity. See Hensher (2009) for a detailed explanation about hypothetical bias.

After deciding about the context, the type of measuring WTP should be selected. WTP can be measured in two ways. One way is to indirectly measure WTP on the basis of consumers' evaluations - can be choices, rankings, or ratings - for several priced product profiles through an evaluation task. The other way is to ask consumers directly to state their WTP for a specific product profile through a question format (Miller et al., 2011). Many studies verify the variation between direct approaches and indirect approaches based on their resulting mean WTP and demand curves (Backhaus et al. 2005; Silva et al., 2007).

Moreover, indirect measurement methods can be further distinguished between compositional, de-compositional and mixed with respect to the way they elicit preferences (Green and Srinivasan, 1990). De-compositional techniques ask the respondents to state their preferences about full product profiles. Compositional techniques ask the respondents to evaluate pairs of individual characteristics that constitute a product profile. Finally, there are mixed approaches which initially ask the respondents to evaluate pairs of separate product characteristics and then to proceed on the evaluation of a small number of product profiles. The different evaluation tasks imply different elicitation procedures for consumer preferences. After these preferences are extracted, WTP can be determined. Green (1984) and Helm et al. (2004) found significant differences between compositional and de-compositional techniques.

Table 1: Classification Table of WTP methods

Alternative Methods For Determining WTP		
<i>Purchase Context</i>	<i>Direct Measurement</i>	<i>Indirect Measurement</i>
<i>Actual WTP</i>	• Market Transactions • Experimental Auctions	• Incentive-Align Surveys
<i>Hypothetical WTP</i>	• Contingent Valuation	• Conjoint (De-compositional) • Self-Explicated Tasks (Compositional) • Adaptive and Hybrid Conjoint (Mixed)

1.2 Motivation of research

With respect to the above discussion about WTP and preference data, the motivation of this thesis is to deepen the knowledge of hypothetical WTP methods by examining how they differ in terms of their mean and demand curves. Particularly, the significant variations reported from previous research regarding the type of measurement and the compositional manner of indirect measurement methods should be further investigated. Each method in table 1 has its own specific design which leads to different kind of questions or tasks for the respondents and eventually to different determination procedures.

Most of the studies in marketing and consumer behaviour literature are examining the differences among actual and hypothetical WTP techniques (Miller et al., 2011; Wertenbroch and Skiera, 2002; Grunert et al., 2009; among many others). But, there is no study to compare more than two hypothetical approaches together in order to illuminate further and improve the way that hypothetical methods provide data input for the determination of WTP and the creation of demand curves. This thesis employs three hypothetical WTP methods in order to conduct a systemic examination of how and in what extent the specific characteristics of each survey method can affect the results.

Conjoint Analysis (CA) (Green and Srinivasan, 1990) is the most prominent methods for acquiring hypothetical preference data and then to determine WTP. Wittink et al., (1994) verify this statement by explaining that because of its superior performance in comparison with other methods, one of the most usual applications of CA is to estimate WTP. CA is located in the indirect type of WTP measurement methods which use a de-compositional procedure for the elicitation of preferences. Contingent Valuation (CV) (Mitsell and Carson 1989) is also a quite common technique for collecting hypothetical preference data. According to Grunert et al. (2009), the method has been especially popular because of its ability to determine WTP when there are no available market prices. CV is located in the direct type of WTP measurement methods.

Paired Comparison-Based Preference Measurement (PCPM) (Scholz et al., 2010) is a recently developed methodology which is also used for the elicitation of consumer preferences in a hypothetical setting, and it is related to the tradition of self-explicated approaches. PCPM, like CA, it is located in the indirect type of WTP measurement methods. However PCPM uses a compositional procedure for the elicitation of preferences. PCPM is an alternate version of Analytic Hierarchical Process (AHP) (Saaty, 1990) which is used to model decision problems.

What distinguishes PCPM is its application to complex products like mobile phones or cars, which are composed of a large number of product characteristics (Wertenbroch and Skiera, 2002). The method has been created in an effort to make the evaluation task easier when the product under examination is constituted by many characteristics which will lead to the creation of large product profiles (that would be the case in CA) difficult to be evaluated by respondents. PCPM has been examined in terms of its validity in comparison with CASEMAP (compositional) and Adaptive Conjoint Analysis (mixed), but it has never been included in WTP studies. Moreover, according to Scholz et al. (2010), the technique results to very precise estimates with respect to consumers' preferences, so it would be of

significant importance to measure its precision in the determination of WTP as compared with other already established methodologies like those of CA and CV. Additionally, there is a lot of methodological interest for investigating how the design of the technique will affect mean WTP and demand curves.

The reasoning behind this research is that hypothetical WTP methods can provide crucial insights to decision makers in situations when time and budget constraints do not allow for the employment of more sophisticated techniques like those used under the actual purchase context. However, little is known about the WTP determination process of hypothetical methods and how this process can affect WTP.

Apparently, the aim of this thesis is to empirically compare CA, CV, and PCPM together through a large-scale research design by analysing their difference regarding their mean WTP and demand curves. This analysis is focusing on the way that technical characteristics of hypothetical methods can affect their performance on determining WTP.

The selection criterion of these methods is that they substantially cover the spectrum of various survey designs and data treatments which are employed for the determination of WTP. This thesis employs several modulations in the design of each method in order to prevent meaningless comparisons and therefore to ensure consistency in the analysis of WTP for the three hypothetical WTP methods.

1.3 Research Questions

For the research objectives of this study the following research questions should be answered:

- 1) Do CA, CV and PCPM differ in terms of their mean WTP and demand curves?
- 2) How the mean WTP and demand curves of CA, CV and PCPM can be affected?
 - a) How the product in question can affect the results?

1.4 Outline

The structure of the remaining sections is unfolded in the following sequence: In Section 2, the process of determining WTP from preference data is explained. In Section 3 a detailed overview of the three methodologies is provided where their theoretical underpinning and examples of their application are discussed. Then in Section 4, a discussion about the current scientific debate of the methods concerning their advantages and disadvantages takes place, but also suggestions from previous studies are discussed. In section 5 the materials and methods used to for the research design, data collection, and determination of WTP methods are explained. In Section 6 the results are discussed. In Section 7 a discussion takes place concerning the results found in the study compared with those exist in the literature. In Section 8 the conclusions drawn from this study are explained followed by recommendations for further research.

2. The process of determining WTP from preference data

A hypothetical WTP study consists of five consecutive stages. Figure 1 represents an overview of these stages. The scope of this section is to give an idea to the reader of the steps that the researcher or the practitioner should take in

order to conduct a hypothetical WTP study and to explain how the actions in each step affect the determination of WTP. Every part of this section will be explicitly explained in the rest of the paper.

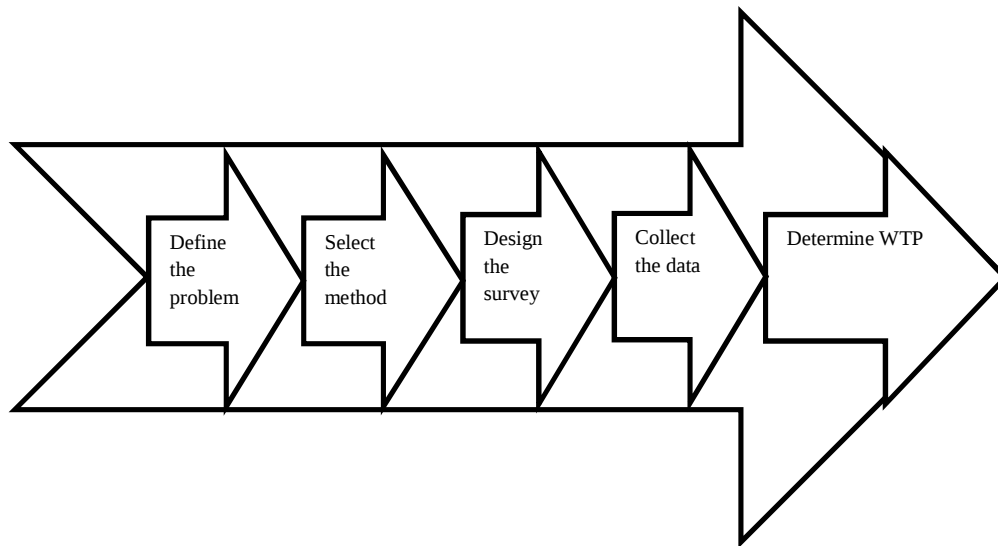


Figure1: An Overview of the stages of a WTP study

2.1 Define the problem

A WTP study starts with a problem that needs to be solved. The problem is how much consumers are willing to pay for a new product concept or a product improvement. So, the first decision that should be taken into consideration in a WTP study it has to do with the definition of the stimulus that respondents have to state their preferences about it. The stimulus refers to a real or a hypothetical product then, according to the available information in the market about the stimulus the relevant method should be selected.

2.2 Select the method

Conjoint analysis is the most successful method for determining WTP. So, this should be the first option for performing a WTP study. Despite its superiority, there are cases that CA is not suitable. CA uses full product profiles for the elicitation of preferences and when the underlying product is comprised by a large number of attributes the profiles to be evaluated will be quite complex. In this study a full profile is constituted by 5 attributes and a partial profile by 3. So, the mental effort of respondents will be increased which may lead to wrong inferences when analysing their responses. Then, Paired Comparisons-Based Preference Method would be the optimal choice. In general, concerning the hypothetical WTP, the indirect approaches outperform the direct ones. However, if the product is completely new which implies that there are no existing prices in the marketplace yet, Contingent Valuation is the recommended alternative. With respect to the selected method the appropriate survey design should be created for the determination of WTP.

2.3 Design the survey

The design of the survey is dependent upon the measurement type of the method, that is direct or indirect, and the compositional manner of indirect methods. Hence according to specific method characteristics the respondents are asked to evaluate a full product profile (CA), pairs of separate product characteristics (PCPM) or a proposed price for a product profile (CV). Hence the creation, presentation, and evaluation task or question format about the stimulus profile differ in each survey design. Also, the way that proposed prices are determined for each stimulus profile differs across the designs. So, there is a specific WTP determination procedure for each survey design. As mentioned in the previous section this study takes seriously into account the differences in each design and conducts the necessary configurations in order to create a common frame for comparing the three methods.

2.4 Collect the data

The sample of the respondents must be representative. The demographics of the sample should match those of the target group. Also the familiarity of the product should be checked as a filter question since respondents who are not sufficiently aware about the product they are not able to provide realistic statements. The data needs to be collected from a specific sized sample. The minimum sample for one method to be testable is 20 respondents. There are two ways of collecting data. One is giving respondents a printed version of the questionnaire and the other is to send them a link of an on-line version. Data collection is a very crucial step, especially in hypothetical methods because they have to approximate real purchase scenarios. So, if the sample is not representative, the data will be of no value, since WTP will be determined from consumers who are not interested for the product in question. Also, if the amount of respondents is not sufficient for analysis the results will be meaningless too.

2.5 Determine WTP

After collecting the data, the WTP for the underlying stimulus can be determined. However, as mentioned above each method implies different determination procedures. The major difference in WTP determination lies on the measurement type. In direct measurement approaches like CV, the WTP can be immediately determined. Specifically, the respondents are being directly questioned if they would pay a price to buy a hypothetical product profile. The price that a respondent accepts for buying a product is the WTP for that respondent.

In indirect measurement approaches the WTP is determined by asking respondents to perform an evaluation task. Nevertheless, in compositional techniques respondents evaluate pairs of separate product characteristics including price. Then, the stated preferences are calculated via these evaluation scores and they are composed to form an overall evaluation for a product profile, can be full or partial. On the other hand, in de-compositional techniques respondents evaluate full product profiles. Then the stated preferences are estimated via these evaluation scores and decomposed to give separate values for each product characteristic. The term product characteristic refers to the attribute levels that used to form a specific product profile. In elicitation preference methods, a product profile is formed by a predetermined number of specific attributes. Also, each attribute is consisted by a limited number of levels.

The common point of both compositional and de-compositional techniques is that both provide scores regarding individual preferences for every attribute level. These scores are called utilities and they are used for the determination of WTP. These utilities are estimated or calculated with regards to the compositional manner of the method.

3. Description of the hypothetical WTP methods in market research

3.1 Conjoint Analysis

Since its introduction to marketing literature by Green and Rao (1971), Conjoint Analysis has become one of the basic hypothetical indirect methodologies in preference studies. According to Green and Srinivasan (1990), CA is considered the most prominent method for acquiring hypothetical preference data. By conducting CA, the researcher can shed light on questions such as: What product attributes are conceived as important or unimportant to the consumer? What levels of product attributes are the most or least desirable in the consumer's mind? What is the share of preference for leading competitors' products versus a company's existing or proposed product (Gustaffson et al., 2007)? How many are the potential market segments which want to buy the specific product? What is price that consumers are willing to pay for a product?

According to Green et al. (2001), CA is considered the favourite methodology of marketers for extracting insights about the way buyers make trade-offs among rival products and suppliers. The major characteristic of CA is that the technique forces the respondent to decide in the same way as he supposedly does in a real market situation, which is by trading off between different alternatives. So, the stated preferences of the respondent for a product offering are implicitly determined by what is important to the respondent. More recent papers have successfully used and expanded further the scope of CA by examining the influence of quality cues on purchase intention (Roest and Rindfleisch, 2010) or the influence of price or country of origin to product quality (Veale and Quester, 2009).

The objective of conjoint analysis is to determine what combination of levels of the different attributes is most influential on decision making. Referring to the definition given by Green and Srinivasan (1990) CA is related to "any de-compositional method that estimates the structure of a consumer's preferences, given his overall evaluations of a set of alternatives that are pre-specified in terms of levels of different attributes". The term 'conjoint', according to mathematical psychologists, is related to the situations where there are measurement scales for both the dependent variable (the ranking or rating of a product profile) and independent variables (the different attribute levels), given the order of the joint impacts of independent variables and a predetermined composition norm (Luce and Tukey, 1964). This means that the evaluation score of the hypothetical profile is decomposed to give scores to the specific attribute levels that were included in the formation the specific profile. Then, it can be predicted for each respondent, which is the most or the least preferred attribute level. The significant amount of between-person variation has driven researchers to conduct CA on individual level in order to preserve efficient estimates and valid statistical inferences (Green and Srinivasan, 1978).

The product profile h ($h=1, H$) which can be described by the vector $i = (i_1, i_2, \dots, i_j, \dots, i_J)$ where i_j represents the level i ($i=1, I$) of the attribute j ($j=1, J$). Then the stated preferences $Y_{h[i_1 i_2 \dots i_j \dots i_J],k}$ of the respondent k ($k=1, K$) for product profile h , where $h[i_1 i_2 \dots i_j \dots i_J]$ connotes the full-profile stimulus h , can be approximated by the additive compensatory model:

$$Y_{h[i_1 i_2 \dots i_j \dots i_J],k} \cong \sum_{j=1}^J x_{i_j k} \quad (\text{Eq. 1})$$

where $\cong$ refers to the least squares approximation or some other kind of fitting process, $\sum_{j=1}^J x_{i_j k}$ denotes the sum of the part-worths for every level of each attribute in profile h , and x is the $i_j k$ part-worth of level i of attribute j for the respondent k . Part-worths show the trade-offs that a consumer makes among different attribute levels.

Therefore, according to Green (1984) the predicted preferences of profile $h[i_1 i_2 \dots i_j \dots i_J]$ for respondent k can be estimated as an additive function of a particular set of attribute levels i_j which is equal with the right hand side of (1):

$$\hat{Y}_{h[i_1 i_2 \dots i_j \dots i_J],k} = \sum_{j=1}^J x_{i_j k} \quad (\text{Eq. 2})$$

In addition, market segmentation can be performed by using the part-worth estimates $x_{i_j k}$ to uncover the potential number of segments and then the description of these segments can be enriched by demographics.

One of the most critical issues concerning the results of CA is the design of the task where respondents will state their preferences. CA is related to the task which respondents are given an amount of various product profiles that have been created through a factorial design of product attributes and attribute levels. Particularly, there are five major alternative procedures which are currently employed to create the evaluation task for respondents:

1. **Full factorial design techniques:** this is the most common approach where the respondent should rate a 0-100 likelihood-of purchase scale for all possible product profiles that can be created by different combinations of attribute levels. This method is used only if there are not too many attributes and attribute levels, because the number of product profiles increases exponentially with them (see the explanation below Figure 2), which can cause information overload and burden the respondent, and therefore prevent the correct answers, leading to measurement and response errors (Green and Srinivasan, 1978).
2. **Fractional factorial main-effects design techniques:** if there are too many attributes, researchers usually employ a fractional factorial approach which can result in a reasonable number of profiles easy to be processed by an individual with average cognitive abilities. These designs emphasize on orthogonal arrays and main effects, whereas they limit or totally exclude any interaction effect.
3. **Hybrid techniques:** each respondent performs a self-explicated evaluation task, which will be explained later in the PCPM section, and then evaluates a subset of the partial product profiles (Green et al., 1981). The resulting utility function is a composite of data obtained from both tasks.
4. **Adaptive conjoint analysis (ACA):** this is a hybrid technique developed by Sawtooth Software (Johnson 1987), where each respondent initially performs a self-explication task and then evaluates a set of partial-profile descriptions, two at a time. These partial profiles usually consist of two or three attributes per stimulus product profile. Then, researchers vary the partial profile descriptions depending upon responses to earlier paired comparisons. The respondent evaluates each pair of partial profiles on a graded, paired comparisons scale. Both tasks are administered by computer (Johnson 1987).
5. **Choice Based Conjoint Analysis (CBCA):** Louviere and Woodworth (1983) came up with an innovative approach named Choice Based Conjoint Analysis (CBCA), in order to expand earlier approaches concerning the ranking or rating of various product descriptions. In their article they provide a breakthrough for CA by merging conjoint and discrete choice modelling approaches (Louviere et al., 2000). CBCA again is performed by the evaluation of product profiles but instead of ranking or rating respondents have to choose the best alternative product profile among a number of choice sets which are usually created via a rotation design approach (Bunch et al., 1996). Furthermore, it is recommended that CBCA should include also a no-choice option in each choice set which gives the right to the respondent to state that he would not choose any of the available profiles (Miller et al., 2011). Elrod et al. (1992), pose that many

papers that have compared the relative performance of the standard CA and CBCA show that both models tend to provide predictions that are equally satisfactory.

In traditional CA, that is factorial designs, the part-worths are majorly estimated by the means of multiple-regression. Then, the regression coefficients reflect the contribution of each attribute level.

Explanation of the ordinary module of Conjoint:

For example, consider the following attributes of a mobile phone and their different levels in Table 2:

Table2. Attributes and attribute levels of a mobile phone

Attributes	Levels			Number of levels
Brand	Nokia	Samsung	Unbranded	3 levels
Memory	32GB	16GB	64GB	3 levels
GPS	Yes	No	-	2 levels
Camera	Yes	No	-	2 levels
Price	€399	€99	€299	3 levels

Now, consider the following hypothetical product profile derived by the above characteristics:

Table3. Hypothetical product profile of a mobile phone

Attributes	Levels
Brand	Nokia
Memory	16GB
GPS	No
Camera	Yes
Price	€299

Also bear in mind that in a full-factorial design the profile in Table 3 can be one of the $3^3 * 2^2 = 108$ possible profiles.

Once part-worths are estimated, predicted preferences of respondent k for profile h can be calculated by Equation 3:

$$\hat{Y}_{h[Nokia, 16GB, GPS_{No}, Camera_{Yes}, Price_{299}], k} = x_{Nokia_{Brand}^k} + x_{16GB_{Memory}^k} + x_{No_{GPS}^k} + x_{Yes_{Camera}^k} + x_{€299_{Price}^k} \quad (\text{Eq. 3})$$

Srinivasan (1982) suggests that price can be one of the attributes in a CA with a finite and predetermined number of levels. Furthermore, if price is one of the attributes included in product profiles generated by a factorial design, it can be estimated by using the same metric as for all the other attributes.

However, Choice Based Conjoint Analysis is nowadays the most common conjoint design used in WTP studies. Therefore, this thesis will employ a CBCA design since choice tasks are frequently used in pricing research (e.g., Orme, 2006; Grunert et al., 2009; Miller et al., 2011). Also, according to Louviere (1988), choice tasks are more realistic than ranking or rating. In choice tasks, respondents are introduced to various product concepts and asked to choose the one they would buy. The rationale of this design is that the evaluation task that the respondents are asked to conduct, it is a very realistic approximation of the actual decision they make in real purchase.

Therefore, the main advantage of CBCA is that it imitates real choice behavior. In CBCA part-worths are considered to be utilities. These utilities reflect the level of satisfaction a person receives from a specific level of the product attribute. According to the principle of monotonicity, higher utility means higher satisfaction. In other words, in real life, people tend adopt a rational choice behavior. So, they try to maximize their total utility when making a choice.

Following the previous notation of traditional CA, in CBCA utility represents the benefit gained when selecting a product profile h which belongs to choice set H . It has to be mentioned that since the respondents are asked to choose product profiles from multiple choice sets, then H belongs to a superset $\mathbb{C}$ which contains all the choice sets in a choice-based conjoint design. Then, the choice set H ($H=1, \mathbb{C}$) can be described by the vector $l = (l_1, l_2, \dots, l_h, \dots, l_H)$ where l_h represents the product profile h ($h=1, H$) which formed by the level i ($i=1, I$) of the attribute j ($j=1, J$). In discrete choice modelling, choices are used for the estimation of utilities.

So, according to Aizaki (2012), the probability that the respondent k ($k=1, K$) selecting alternative l_h from choice set H can be formulated by the general logit form:

$$P(l_h) = \exp(V_{l_h}) / \sum_{h=1}^H (V_{l_h}) \quad (\text{Eq. 4})$$

In addition, V_{l_h} is a systematic component of utility that is assumed to be a linear additive function of the independent variables $x_{i,j,k}$ which at the end can be formulated in a similar way with that of Equation 2:

$$V_{l_h}^{H_{[i_1 i_2 \dots i_j \dots I_j]k}} = \sum_{j=1}^J x_{i_j k}^H \quad (\text{Eq. 5})$$

This study treats WTP as the additional amount of money that a respondent is willing to pay for an improved version, compared to the base version of a product. The willingness to pay can then be simply determined as a ration between the absolute difference of the utilities of the attribute levels in the improved and base version divided by the utility of price. Moreover, Miller et al. (2011), Wertenbroch and Skiera (2002), and Balisteri et al. (2001) among others, have shown how to create demand curves from WTP estimates by survival tables.

3.2 Paired Comparison-Based Preference Measurement

Paired Comparison-Based Preference Measurement is an indirect measurement approach. PCPM has been proven to be a prudent methodology for measuring consumers' preferences created by Scholz et al. (2010). Its rationale is similar with that of CA, since it considers a pre-specified number of attributes each with a certain number of levels for the evaluation of a product profile. However, the method is located near to the methodological stream of self-explicated tasks. So, PCPM uses a compositional manner to elicit consumer preferences.

Self-explicated data collection reflects situations where each of the respondents has to rate the desirability of each choice set of attribute levels on a 0 to 100 scale and in turn to rate the attributes on an importance scale (Srinivasan, 1988). The choice set of attribute levels is formed by paired comparisons which means that all the levels that belong in one attribute are presented in pairs where the respondent has to state which level among the two is the most desirable according to his preferences. The same procedure takes place between the attributes themselves where the respondent has to state which attribute is the most important for him between the two. The respondent has to evaluate all the possible paired comparisons among the attributes and their levels.

According to Scholz et al. (2010), the method is an alternate version of the Analytic Hierarchical Process (AHP) (Saaty, 1990). AHP has made a substantial contribution in business research and particularly in managerial decision making (Forman and Gass, 2001; Vargas, 2006). Concerning the application of AHP in consumer research it has been evident that the technique has high predictive accuracy with respect to preference measurement (Helm et al., 2004a; Helm et al., 2004b; Mulye, 1998). Specifically, Mulye (1998) indicates that the most intrinsic feature of AHP in the field of consumer behaviour lies in its effective elicitation of preferences for complex products, that is products with several attributes and attribute levels.

Vaidya and Kumar (2006) describe AHP as a multiple criteria decision making tool. The rationale behind the technique is to hierarchically divide a decision making problem into less difficult sub-problems. Then, each of these sub-problems constitutes a certain number of paired comparisons. The major aim of AHP is to reduce the cognitive burden of the respondent that is expected to increase, for example in the CA task, when the number of attributes and their levels is high enough to lead to the creation of numerous and complex product profiles, no matter what data collection design is employed (like fractional factorial main-effects design).

It has to be indicated though, that the applicability of AHP is not limited to complex product profiles. For instance, Helm et al. (2003) conducted an AHP study for students' preferences about universities which included 6 attributes, specifically 5 with 3 levels and 1 with 2 levels. Despite of its promising results, Meißner and Decker (2009) indicate that AHP has not gained significant interest in the field of marketing research so far. Scholz et al. (2010) made a successful attempt to bring AHP closer to the demands of marketing and consumer decision making. PCPM is a task in which respondents are asked to make paired comparisons. These paired comparisons represent the trade-off that a consumer has to make among the attributes and their corresponding levels.

In Figure 2 a pictorial representation of PCPM, regarding the hierarchical division of a product evaluation problem in smaller sub-problems, it is exhibited. PCPM employs a three layer hierarchy in order to decompose the product in terms of its attributes and their corresponding levels that determine its total utility and related preferability (Meißner et al., 2011).

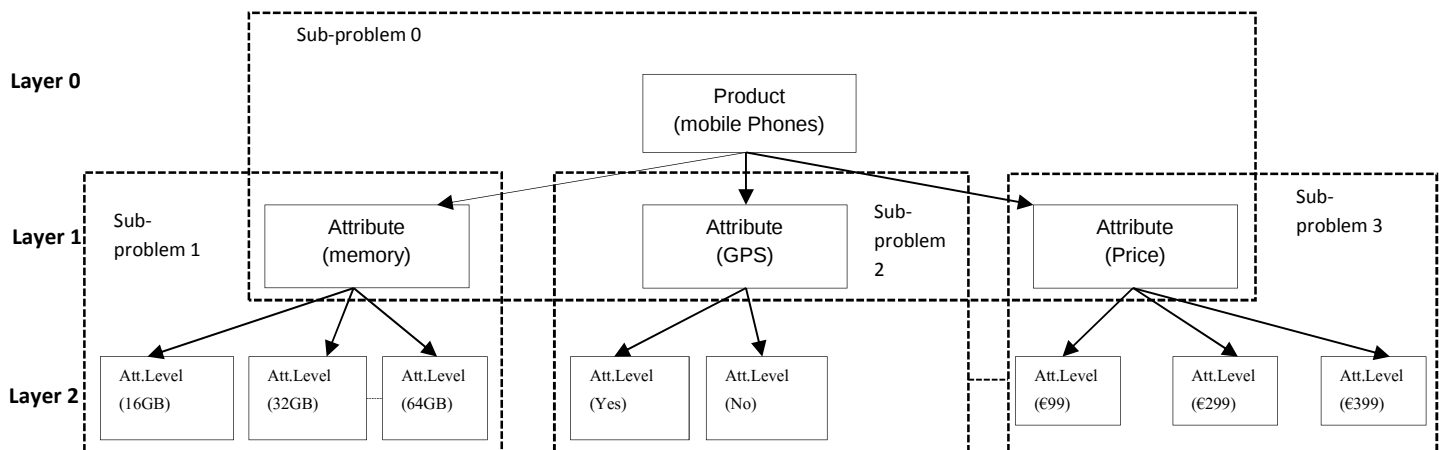


Figure 2. Hierarchical Structuring of a product evaluation problem (Scholz et al, 2010)

A specific sub-problem contains all paired comparisons of the elements in a particular layer in the hierarchy. By looking at Figure 2, it can be seen that PCPM describes a mobile phone as a concrete product at Layer 0, its attributes at Layer 1, and their corresponding levels at Layer 2. Analytically, Sub-problem 0 contains the paired comparisons among the attributes (memory, GPS, price), and the respondent should state which of the two compared attributes prefers more by assigning different importance scores. For the rest of the sub-problems the respondent should state which of the two attribute levels of an attribute prefers more by assigning different desirability scores. Specifically, Sub-problem 1 contains the paired comparisons among the attribute levels (16GB, 32GB, and 64GB) that constitute the memory of a mobile phone. Sub-problem 2 entails a paired comparison among the inclusion or exclusion (Yes and No) of a GPS software in mobile phone. Sub-problem 3 contains the paired comparisons among the attribute levels (€99, €299, and €399) that describe the price range of a mobile phone. For more detailed description of the method read the work of Scholz et al. (2010).

Moreover, Meißner and Decker (2009), with regards to hierarchical approaches, suggest a bottom up evaluation of the hierarchy in order to make respondents cognizant about the attribute levels and their ranges before they proceed to the evaluation of the whole attributes. In other words, the respondents should initially evaluate all pairs of attribute levels at the bottom level and then continue with the pairwise comparisons of the attributes at the higher level (Figure 3).

A: Paired Comparison of Attribute Levels

Which capacity of memory do you prefer when do you prefer when thinking about a mobile phone?										
16GB	Absolutely prefer left	Strongly prefer left	Considerably prefer left	Weakly prefer left	Indifferent	Weakly prefer right	Considerably prefer right	Strongly prefer right	Absolutely prefer right	32GB
	1	2	3	4	5	6	7	8	9	

B: Paired Comparison of Attributes

Which of the two features, GPS or memory, is more important to you when purchasing a mobile phone?										
GPS (With, Without)	Left absolutely more important	Left strongly more important	Left considerably more important	Left weakly more important	Indifferent	Right weakly more important	Right considerably more important	Right strongly more important	Right absolutely more important	Memory (16GB, 32GB, 64GB)
	1	2	3	4	5	6	7	8	9	

Figure 3. Paired Comparisons for Attribute Levels and Attributes

The stated paired comparisons for a sub-problem can be depicted in a paired comparison matrix. An example of such matrix, concerning the attributes used in the mobile phone scenario, is presented in Table 4.

Table 4. Paired comparison matrix for mobile phone attributes

	Brand	Memory	GPS	Camera	Price
Brand					
Memory					
GPS					
Camera					
Price					

The number of paired comparison increases exponentially with the number of elements (attributes or attribute levels) in a sub-problem. Totally $n*(n-1)/2$ paired comparisons for the n ($n=1, N$) elements of every sub-problem should be conducted by the individual respondent to form the aforementioned matrix. Particularly, the respondent has to evaluate $J*(J-1)/2$ paired comparisons for attribute importances in Sub-problem 0, plus $I_j * (I_j - 1)$ paired comparisons for attribute level desirabilities in each of the remaining sub-problems. The mobile phone that has been used so far to provide a practical example regarding the application of CA is composed of 5 attributes. Therefore, $5*(5-1)/2 = 10$ paired comparisons should be done for the estimation of the importance weight of each attribute. The below diagonal elements (shaded area) represent the preferences for the row attributes over the column attributes.

The researcher or the practitioner when they collect paired comparison data they can move on to the utility derivation for the profile in question. According to Green (1984), the self-explicated utility U of the hypothetical product profile h ($h = 1, H$) for the respondent k ($k = 1, K$) is calculated through the desirability scores u_{ijk} of the attribute levels, and importance weights w_{jk} by the weighted additive model:

$$U_{h[i_1 i_2 \dots i_j \dots I_j],k} = \sum_{j=1}^J w_{jk} u_{ijk} \quad (\text{Eq. 6})$$

where, as explained in the CA section, the subscript h denotes the specific set of levels which are dummy variables, that are included in the particular product profile.

The major difference between PCPM and CA is that the former is compositional in the sense that calculates U for the profile h as a weighted sum of importance weights and related desirability scores as separated and explicitly judged by the consumer, whereas the latter is de-compositional since it starts with the respondent's overall evaluations for different product profiles which are then used to explicate the values given to the attribute levels (Green and Srinivasan, 1978).

In PCPM the predicted preferences $\hat{Y}$ of the respondent k are equal with an additive function of the products of importance weights and desirability scores. Also, $\hat{Y}$ it is assumed to be quite same with that of conjoint depending on a scale factor. $\hat{Y}$ of profile h for respondent k is given by the formula:

$$\hat{Y}_{h[i_1 i_2 \dots i_j \dots I_j],k} = a_k + b_k U_{h[i_1 i_2 \dots i_j \dots I_j],k} = \sum_{j=1}^J x_{ijk} \quad (\text{Eq. 7})$$

WTP and demand curves can be determined in analogous ways to those used in CBCA if price is one of the product attributes.

The most important issue regarding the design of the task in PCPM is the Bipolar scale where the respondent are asked to state their preferences. Paired comparison methodologies gather subjective information and create relative importance and relative desirability information.

This kind of information can be interpreted as constant ratios (Saaty, 1980) or as constant differences (Turner, 1996). The PCPM scale measures at the same time the direction and the strength of the respondent's preferences

(Hartvigsen 2003). Figure 4 depicts the ratio scale which is employed by Scholz et al. (2010). In particular, there is a continuum of 9 scale levels and each level represents corresponds to a specific ratio value for respondent's preferences. However it should be noted that ratio data which is derived from relative judgments among two attributes or attribute levels should not be confused with ratio data that corresponds to interval data with a fixed zero point (Turner, 1996).

Scale level	Verbal statement	Preference value
1	16GB is absolutely preferred to 32GB	9.00
2	16GB is strongly preferred to 32GB	5.20
3	16GB is considerably preferred to 32GB	3.00
4	16GB is weakly preferred to 32GB	1.73
5	16GB and 32GB are equal	1.00
6	32GB is weakly preferred to 16GB	1/1.73
7	32GB is considerably preferred to 16GB	1/3.00
8	32GB is strongly preferred to 16GB	1/5.20
9	32GB is absolutely preferred to 16GB	1/9.00

Figure 4. PCPM Constant Ratio Scale for preference measurement

In the Constant Ratio Scale the ratios S_{q+1} / S_q are evenly dispersed. The scale provides the geometric increase or decrease among the adjacent scale levels, i.e. $S_{q+1} / S_q = \text{constant}$ for, where q is the scale level, $q = 1, \dots, 8$. The PCPM scale values can be also interpreted as the exponential decrease or decrease of the number 3: $3^2, 3^{3/2}, 3^1, 3^{1/2}, 3^0, 3^{-1/2}, 3^{-1}, 3^{-3/2}, 3^{-2}$. The exponential decrease or decrease of the number 3 is used because it makes the scales perfectly symmetric, which means that every scale level is way 1.73 points from the other.

The reason for the modification of the scale is the minimization of the error variance in order to increase the validity of the scale. In other words, the rationale of this alteration is to prevent for the Type III measurement error (Meißner et al., 2011). Type III error might result from the use of ratio scales in empirical applications. According to Hurley (2001) respondents are not sure about the precision of the ratio between two elements. The major idea of type III error is that respondents are expected to be discriminating towards the middle-point of a constant differences scale, so error the variance will be low. However, in constant ratio scale the respondents are expected to overestimate the respective preference ratio by using the end-points of the scale, hence the error variance will be high (Meißner et al., 2011).

Apparently, Figure 5 depicts the interval scale which is used in the study of Scholz et al. (2010), where each level refers to a particular interval value for respondent's preferences.

Scale level	Verbal statement	Preference value
1	16GB is absolutely preferred to 32GB	4
2	16GB is absolutely preferred to 32GB	3
3	16GB is considerably preferred to 32GB	2
4	16GB is weakly preferred to 32GB	1
5	16GB and 32GB are equal	0
6	32GB is weakly preferred to 16GB	-1
7	32GB is considerably preferred to 16GB	-2
8	32GB is strongly preferred to 16GB	-3
9	32GB is absolutely preferred to 16GB	-4

Figure5. PCPM Constant Differences (Interval) Scale for preference measurement

The difference between ratio and interval data derived from relative judgements is in the paired comparison matrix. A paired comparison matrix has the general form of:

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_{11} & \cdots & \mathbf{a}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{a}_{n1} & \cdots & \mathbf{a}_{nn} \end{pmatrix}$$

where $\mathbf{a}_{ij}$ represents the preference value obtained from the paired comparison between element i and element j . According to Scholz et al. (2010), the elements $\mathbf{a}_{ij}$ and $\mathbf{a}_{ji}$ are symmetric (see Figures 6 and 7). The main diagonal elements always take the value of 0 for the constant differences scale and 1 for the constant ratio scale, and both 0 and 1 they refer to indifference in respondent's preferences because an attribute cannot be less or more preferred than itself.

Moreover, it has to be indicated that the reference to respondent k is avoided in order to simplify notation and save space (otherwise paired comparison data from each respondent should be plugged into K different matrices specific to N sub-problems).

The major distinction between interval and ratio scales lies in the cell values of $\mathbf{A}$. The preference values of a pairwise comparison are related with an additive function (for the interval scale) or a multiplicative function (for the ratio scale), of the calculated importance weights (and desirability scores).

For interval scales the paired comparison matrix of importance weights takes the form of:

$$\mathbf{A}^{\mathbf{w}_{\text{interval}}} = \begin{pmatrix} \mathbf{0} & \cdots & \mathbf{w}_1 - \mathbf{w}_j \\ \vdots & \ddots & \vdots \\ \mathbf{w}_j - \mathbf{w}_1 & \cdots & \mathbf{0} \end{pmatrix}$$

where a paired comparison $\mathbf{a}_{ij}$ measures the difference in importance weights (or desirability scores) between two elements, i.e. $\mathbf{a}_{ij} = \mathbf{w}_i - \mathbf{w}_j$.

Similarly, for desirability scores the paired comparison matrix takes the form of:

$$\mathbf{A}_j^{\mathbf{u}_{\text{interval}}} = \begin{pmatrix} \mathbf{0} & \cdots & \mathbf{u}_{1j} - \mathbf{u}_{lj} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{lj} - \mathbf{u}_{1j} & \cdots & \mathbf{0} \end{pmatrix}$$

On the other hand, regarding ratio scales, the paired comparison matrix takes the form of:

$$A^w_{\text{ratio}} = \begin{pmatrix} 1 & \cdots & w_1/w_J \\ \vdots & \ddots & \vdots \\ w_J/w_1 & \cdots & 1 \end{pmatrix}$$

where a paired comparison a_{ij} measures the ratio of the importance weights of two elements, i.e. $a_{ij} = w_i/w_j$.

Similarly, for desirability scores the paired comparison matrix takes the form of:

$$A^u_{\text{ratio}} = \begin{pmatrix} 1 & \cdots & u_{1j}/u_{Ij} \\ \vdots & \ddots & \vdots \\ u_{Ij}/u_{1j} & \cdots & 1 \end{pmatrix}$$

For further explanation regarding the derivation of importances and desirabilities see the technical appendix of Scholz et al. (2010) and Hartvigsen (2003). Also, the elements of both difference and ratio matrices have another important relationship, since $\log(w_1/w_j) = \log w_1 - \log w_j$ (the same holds for desirabilities). This is an evidence of the convergent validity of the two scales. Also, Scholz et al. (2010) found that there is no significance difference on the individual hit rates between the ratio and interval scales. This is an evidence of the predictive validity of the scales. Additionally, the results of Scholz et al. (2010) vouch those of Hauser and Shugan (1980) which means that one cannot single out one of the scales as more appropriate. This thesis employs ratio scales.

3.3 Contingent Valuation

Contingent Valuation is a direct measurement WTP method and has been particularly known for determining WTP where market prices are not available in the market. This method provides respondents the chance to decide in economic terms the valuation of a non-market good (Grunert et al., 2009). As being defined by Mitchell and Carson (1989), CV is a method that “uses survey questions to elicit people’s preferences for public goods by finding out what they would be willing to pay for specified improvements of them”.

The most common use for CV is to extract respondents’ WTP for alterations in the formation or management of environmental assets. Arrow et al. (1993) indicate that according to NOAA [*National Oceanic and Atmospheric Administration*] guidelines a typical CV study aims to uncover the WTP of citizens for the preservation of natural resources. CV provides respondents with information about the specific actions that are planned to take place by the government or a non-profit organization and then asks them whether they would sacrifice an amount of money to support these actions.

CV because of its advantage to elicit WTP where there is no market data available or no tangible products to be tested, it has been expanded from a natural resources’ valuation methodology to a marketing instrument. Although the technique is not so popular like CA, it has been employed for obtaining WTP estimates for market goods like food products (Boccaletti and Nardella, 2000; Grunert et al., 2009; Gil et al., 2000).

The respondent’s task starts with a hypothetical scenario about the improvement of a product. Respondents are usually given specific information about the product and then they confronted with a question or questions that concern their WTP for buying the improved version. So, from a CA perspective, respondents are asked to compare

two product profiles. However, CV is focusing on a few important attributes of a product (Adamowicz et al, 1998). So instead of explicitly introducing all the five attributes within a complete product bundle, CV design is focusing only on the core attributes of the product that are going to be improved. According to Mitchell and Carson (1989) the other name of CV is Two Alternative Method, which means that respondents are asked to answer a hypothetical question about how much extra money they would be willing to pay in order to buy an improved version (product profile 2) compared to the cost of the base version (product profile 1).

So, the scenario that the respondents will have to read in a CV task would be like:

“Consider a Nokia mobile phone with 16GB memory, without GPS and cost of \$299. Would you pay €100 more in order to get an improved version of this product with GPS and 32GB memory?”

There are three ways to perform a CV study concerning the data collection process:

- 1) **Open-ended questions:** where the respondent is asked to state the amount of money she is willing to pay.
- 2) **Sequential bids:** where the respondent is asked whether or not he would pay or accept a specific price. Concerning this approach, there are continuous repetitions of the question regarding WTP using higher and lower amounts with respect to the primal response until the true value is eventually bracketed.
- 3) **Close-ended questions:** where the respondent is asked, via a dichotomous question, whether or not he would pay or accept a single price.

This study employs the method of sequential bids, where the WTP is determined as the price that a respondent accepts in the highest bid. CV results can also be modelled in a logit form but this is beyond the scope of the study.

4. Scientific debate for the hypothetical WTP methods

4.1 Conjoint Analysis

According to Grunert et al. (2009), the use of CA for hypothetical WTP determination has received severe criticism because the evaluations or choices of the respondents do not have any budget constraint. Therefore, respondents are not committed to anything which leads to the overestimation of WTP. Another argument against the validity of CA in the measurement of WTP is the systematic comparison of different product profiles which are all described by the same attributes, which may not reflect upon all types of consumer-decision making in real life.

Although hypothetical WTP methods are vulnerable to hypothetical bias, may still drive to the proper demand curves and pricing decisions. Miller et al. (2011) stress that methods that create biased WTP should not be entirely avoided. On the other hand, a hypothetical approach is capable of predicting quite accurate demand curves, and thus leading to pricing decisions that are not significantly different from those derived from actual WTP techniques. Another important insight than can be extracted from the work of Miller et al. (2011) is that hypothetical WTP methods may provide better results than actual ones, with respect to the category that the product belongs, since consumers may use different cognitive processes to state their preferences depending whether the product is cheap or expensive.

4.2 Contingent Valuation

Similarly to CA, the validity of CV has been questioned for the same reasons like that of hypothetical bias that lead to overestimation of WTP. So, there is much of scepticism against the performance of CV. For instance, according to Arrow et al. (1993), the detractors of CV argue that respondents do not act in a rational manner since they fail to

take the questions of the survey under serious consideration because the results are not binding. However, Duffield and Patterson (1991) suggest that the deviation of hypothetical direct WTP from actual direct WTP is small enough and predictable enough that WTP scores could be discounted for potential overstatement and used in continue as conservative values of WTP. Additionally, the arguments in favour of the usefulness of CA as being supported from the relevant studies of Miller et al. (2011) and Grunert et al. (2009) stand also for CV as a hypothetical method.

4.3 Paired Comparison Based Preference Measurement

The major reason for the development of PCPM is to be used as a remedy for the shortcomings of self-explicated tasks. Scholz et al. (2010) found that PCPM yields more accurate results in comparison to Adaptive Conjoint Analysis (ACA) and CASEMAP (Computer Assisted Self-Explication of Multi-attributed Preferences) in terms of consumers' preferences. However, the study took place in the context of high involvement products. A way to shed more light in the strengths and weaknesses of PCPM is to apply it also in low-involvement products. Furthermore, it would be interesting to test the performance of the technique in the elicitation of preferences and finally the estimation of WTP when compared with other modules of CA, for example with that of CBCA, but also with other of methods like that of CV.

5. Materials and Methods

5.1 Stimulus

In an effort to uncover any differences in WTP estimates and the shapes of demand curves that can be affected by different product categories, two products (stimuli) were used. The first product was a mobile phone representing the expensive category. The second product was a computer PC mouse representing the low cost category. The two products were unbranded since brand can significantly affect WTP (Steenkamp and Heerde, 2010) and the examination of this effect was not a part of this study. The attributes and their corresponding levels for each product were chosen by literature retrieval for the creation of realistic and comparable profiles with those that exist in the marketplace. Table describes the attributes and their levels that employed method. Actually, Table 5 is an augmented representation of all the elements (prices, attributes and levels among others) which were employed for this research. Each of them is introduced and discussed later in this Section.

In Section 2 it was explained that when one wants to compare the results of different methods at once, he should adopt a methodological procedure which will ensure the comparability between those results. Regarding the particular study this procedure can be described by two steps. The first step was to focus only in a single improvement of two levels of two attributes included in all the three methods. The first levels of the two attributes reflected the base version of the product in question and the immediate next levels reflected the improved version. The second part of this methodology is related with the acceptable range of WTP values in each method, and will be discussed later in the section of empirical comparison of WTP methods where its explanation will be more valuable. The next paragraph explains the particular selection of two attributes and the implications that this methodology had in relation to the representation of the stimulus in each of WTP methods.

As mentioned in Section 3, a hypothetic scenario about a product improvement is given to the respondents in a Contingent Valuation design for the determination of Willingness to Pay. In CV only two attributes should be used to form the stimulus in the design, plus the price. Furthermore, the levels of the attributes should be also two, the first levels to represent the base version and the second levels the improved version (see Table 5). So, in order to attain comparable results, only these two attributes were also used for the determination of WTP in the other two methods. However, in CBCA more attributes were included in the design. Initially, additional attributes were needed because the hypothetical full product profiles could not be realistic with only two attributes. The additional attributes

appeared only as confounders and were not taken into account for the determination of WTP. On the other hand, in Paired Comparison-Based Preference Method only these two attributes, plus the price were employed in the study since the design of the particular technique allowed for similar formation with CV without harming the results. However, in PCPM additional attribute levels were used, again as confounders, in order to satisfy a meaningful number of paired comparisons. The monetary values of the price levels they were defined differently in each method according to specific design requirements (see table 5). A detailed explanation regarding the experimental design of each WTP method for both products is given in the remainder of this section.

The improvement of PC mouse contained the inclusion of two extra attributes (these attributes were comprised by two levels - yes/no). In order to ensure the understanding of the respondents concerning the contribution of these attributes to the functionality of the product, the questionnaires created for PC mouse presented also pictures and written explanations (see appendix A). Also, at the end of the survey respondents were asked if they clearly understood the advantages of these additional attributes.

Table5. Levels and attributes that used in the current study

Mobile phone	Computer Mouse
Attributes used for the determination of WTP	
Camera (CBCA, CV, PCPM): 5.0 Mp (€40) / 10.0 Mp (€80) / 12.0Mp (€120) – The first two levels were used for WTP determination	Extra* Buttons (CBCA CV, PCPM): Yes (€10) / No (€0)
Memory(CBCA,CV, PCPM): 16GB (€50) / 32GB (€100) / 64GB (€150) – The first two levels were used for WTP determination	Air Motion**(CBCA CV, PCPM): Yes (€20) / No (€0)
Blocks in CBCA	
Additional Blocking Factor: 5 Groups	Additional Blocking Factor: 4 Groups
Confounders in CBCA	
Colour: Black/White/Other, Not priced	Ergonomic Design: Yes/No Not Priced
Processor (CBCA): 1.0MHz (€30) / 1.3GHz (€60) / 1.7GHz (€90) , Detractor	2 Years Guarantee (CBCA): Yes (€10) / No (€0), Detractor
Compound attributes in CBCA	
MEM-CAM: 5.0 Mp/16GB, 10.0 Mp/32GB, 12.0Mp/64GB, 10.0/ Mp64GB, 12.0Mp	AIR-BUTTONS: No/No, Yes/No, No/Yes, Yes/Yes
Price stimuli	
Additional reservation Price levels (CBCA): €0, €20, €40, €60, €80 (scale factor of €20)	Additional reservation Price levels (CBCA): €2, €4, €6, €8 (scale factor of €2)
Price of basic version (CBCA): €350	Price of basic version (CBCA): €15
Price Levels (PCPM): €350, €450, €550, €650	Price Levels (PCPM): €15, €30, €45, €60
Initial Bids (CV): €50, €75, €100, €125, €150	Initial Bids (CV): €5, €10, €15, €20, €25

*The term ‘extra’ refers to additional buttons except the right, the left, and the scroll wheel. These keys can be used for performing additional PC functions from the mouse, without using the keyboard, like volume setting, internet surfing, or gaming among many others.

**With the Air Motion one can handle spreadsheets, word documents, and presentations with natural hand movements. Available at: <http://www.gyration.com/products/air-mouse-mobile>.

5.2 Study Design

5.2.1 Choice Based Conjoint Analysis:

As mentioned in Section 2, that is the description of WTP’s determination process via the use of preference data, CBCA was used as the optimal conjoint module for a WTP study. So, for both products in question, the respondents chose their mostly preferred hypothetical product profile from a number of choice sets that were created through a computer-generated design (Louviere and Woodworth 1983). The number of choice sets and the number of hypothetical profiles - plus a none option for each choice set within each choice set - were created via the use of the R software by employing orthogonal main-effects and rotation designs, according to the attributes and their levels employed to form each of the two products. See Aizaki (2012) and Aizaki and Nishimura (2008) for detailed description about the creation of efficient designs of Choice experiments in R.

The total price for each product profile was determined as a linear function of a constant amount of money which represented the base version which is a product concept with no improved features, plus the individual prices of any additional or improved feature included in the specific profile. The base price and the individual prices were selected in an effort to form representable total profile prices with respect to those that exist in the market (see Table 5). In addition, because consumers have different reservation prices (referring to their real WTP, or the maximum amount they are willing to pay), due to their idiosyncratic nature and income (Kohli and Mahajan, 2001; Jedidi and Zhang, 2002), an extra attribute was included to the formation of the product profiles with its levels representing a scaled increase (see Table 5) of additional prices relative to their total price.

The total price of the hypothetical profile h is being defined by the formula:

$$TP_h = BP + \sum_{j=1}^J c_{ij} + RP_i \quad (\text{Eq. 8}),$$

where BP represents the base price, c_{ij} is the cost of attribute level i which belongs to attribute j and RP_i is the additional reservation price level which assigned to this profile via efficient design implementation as every other attribute level.

So, according to Table 5, a mobile phone profile with white colour (colour attribute was not priced but was included in the study to in an effort to ensure a satisfactory product description), 1.3GHz processor (€60), camera 10Mp (€80), memory 64GB (€150) and an €80 reservation price (fifth level) has a total price of $350+290+80=720$ euros.

The fieldwork took place among two nationalities. Also, the total amount of choice sets across respondents was split by a blocking factor (Ryan and Morgan, 2007) which was included as an additional attribute in the design. For each block (group) two different versions with varied order of the attributes in the profiles were disseminated.

The rationale for taking into account the reservation prices, cultural context, and attribute order in the profiles was to minimize measurement errors. Another issue was taking into account for preventing this type of errors. It has been explained in the Section 3 that Contingent Valuation uses only one or two core product attributes for the elicitation of a person's preferences regarding the improvement of their levels or the addition of an extra feature. This means CV takes into account the interaction effect of the two attributes used in the study because they are presented together as one single improvement in the given scenario. This implies that the improvement of one attribute is significantly related with the improvement of the other. However, the efficient design used in this study ignores any interaction effects between attribute levels and this would cause a problem in the comparison of the estimates of the two methods. Efficient designs are used to limit the total amount of profiles. In such designs there is not sufficient number of profiles to ensure all possible combinations of attribute levels. Therefore interaction effects cannot be captured.

For the CBCA a compound attribute (see Table 5) was created by the combination of the corresponding levels of the two attributes which were used for the WTP estimation. Particularly, the levels of the two attributes used for the determination of WTP were combined in a predetermined way to ensure that the base and improved version of mobile phone profiles would be present in every choice set.

5.2.2 Paired Comparison-Based Preference Methodology:

The number of paired comparisons that respondents had to state their desirability and importance scores was defined according to the number attributes and their levels of each product presented in Table 5, considering the formulas explained in Section 3. The price levels were defined in order to be comparable with the total prices in the profiles of CBCA. The same attributes that were used in CBCA to form the compound attribute were used in PCPM. However, because in PCPM the attribute levels did not presented as combined to the respondents the method can capture only main effects. Unfortunately, the nature of the method does not allow for any modulation regarding interaction effects, since it would be meaningless to compare this compound attribute with a single one.

5.2.3 Contingent Valuation:

The respondents were given a scenario regarding the improvement of the product concept in order to state their WTP. Apparently, they were informed about the current formation of the product (base version) and then they were asked to state, in a dichotomous question format, if they were willing to pay or not for a change (improved version) in the same two attributes included in the other two methods. In sequential bids the starting value should vary (Krarup and Russel, 2005). So, total amount of starting values was five per product. The starting values were randomly assigned. The follow up bids were 50% more or less than the initial bid with respect to the positive or negative reaction of the respondent in each bid. Again the initial bid values were chosen in an effort to make comparable price levels with those of the other two methods (see Table 5).

5.3 Data Collection

The data were collected via the means of surveys. Initially, there were six different experimental groups (3 methods and 2 products) developed in a between subjects design. The participants of each group were randomly selected and they were independent from those in the other groups. The overall sample of the study was comprised by 100 for the mobile phone and 99 for PC mouse, randomly selected Greek and Dutch respondents, males and females, 18 to 60 years old. Particularly, for mobile phone the respondents were 56 (CBCA), 24 (CV), and 20 (PCPM). Similarly, for the PC mouse, the respondents were 20 (PCPM), 24 (CV), and 55 (CBCA). The initial versions furtherly divided into more sub-versions according to nationalities, blocks, attribute order, initial bids, and gender. In table 6 information about major demographic characteristics for the respondents participated in the study are provided.

Table6: Description of the major demographic information in mobile phone study.

	CBCA	CV	PCPM
Gender (percentages)	50%males, 50%females	75%males, 25%females	70%males, 30%females
Nation (percentages)	54%Greeks, 46%Dutch	54%Greeks, 46%Dutch	75%Greek, 25%Dutch
Age (mean, SD)	27.54, 3.98	25.42, 4.82	26.7, 6.19
Income (mean, SD)	9642.56, 9056.23	8332.83, 5527.7	6995.5, 5099.02

Table7: Description of the major demographic information in PC mouse study.

	CBCA	CV	PCPM
Gender (percentages)	64%males, 36%females	46%males, 54%females	80%males, 20%females
Nation (percentages)	59%Greeks, 41%Dutch	75%Greeks, 25%Dutch	60%Greek, 40%Dutch
Age (mean, SD)	27.24, 6.44	26.68, 3.34	28.7, 5.63
Income (mean, SD)	10592.72, 8491.51	7917.17, 4545.3	5999.5, 3000

The respondents were graduate, undergraduate, PhD students and university professors, but also professionals and entrepreneurs. The fieldwork took place in Wageningen University, City of Wageningen, and Internet. Each survey started with a filter question asking the familiarity of the respondents. Respondents with low familiarity were excluded from the study. The surveys designed in such a way to be completed in less than 15 minutes in order to reduce respondents' fatigue.

5.3 Determination of mean WTP and creation of demand curves

5.3.1 Choice Based Conjoint Analysis

Madansky (1980) stated first that conjoint analysts could use the random utility framework. Hence, CBCA implies more advance estimation techniques like that of the maximum likelihood and not that of the ordinary least squares. According to Louviere et al. (2000), almost all CA techniques can be viewed as a special case of the Random Utility Theory (RUT) paradigm that reflects a generalized framework for the understanding and modelling of human behaviour, and particularly to the purpose of this paper preferences and choices. The RUT implies the use of discrete choice modelling. The Multinomial Logit (MNL) is the most common discrete choice model. However, MLN focuses on the individual as the unit of analysis and considers the individual's characteristics as explanatory variables. On the contrary, Conditional Logit (CL) focuses on the set of alternatives for each individual and the explanatory variables are characteristics of those alternatives.

Since in this thesis the characteristics were varied across individuals the CL was the recommended model to be used. However even CL was not appropriate since the comparison of methods was conducted on individual level and CL gives estimates for the overall sample. So, the Arora and Huber's (2001) hierarchical Bayes (HB) logit, by using the R software, was finally employed to estimate individual utilities. Nevertheless, the CL model was also used to attain a general view of the CBCA sample regarding the effects of each attribute level included in the creation of product profiles. The CL logit was used to check for the effects of blocking factor, reservation prices, gender, income, age, nationality, order of attributes, and familiarity on the responses. These effects were not significant meaning that the stated preferences of responses were not affected by other factors than the stimulus.

The individual WTPs were calculated in Excel software for each individual using a piecewise linear interpolation in order to mitigate the problem of fat tails that appears in CBCA (Miller et al, 2011). Simple linear interpolation usually results in distributions with fat tails which cannot predict the portion of respondents with high WTP values (see Appendix B). Particularly, a similar procedure with that introduced in the paper of The Jedidi and Zhang (2002), it was employed because of the inclusion of different reservation prices in product profiles. The procedure used to calculate individual WTPs can be explained as:

According to the individual utility scores derived by HB logit,

- If $0 < (U_I - U_B) < (U_{RP_0} - U_{RP_1})$, then $WTP = P_C + a \frac{U_I - U_B}{U_{RP_0} - U_{RP_1}}$ (Eq. 9)

- If $(U_{RP_0} - U_{RP_1}) < (U_I - U_B) < (U_{RP_0} - U_{RP_2})$, then $WTP = P_C + RP_1 + a \frac{(U_I - U_B) - (U_{RP_0} - U_{RP_1})}{U_{RP_2} - U_{RP_1}}$ (Eq. 10)

however,

- If $0 > (U_B - U_I) > (U_{RP_0} - U_{RP_1})$, then $WTP = P_C - a \frac{|U_B - U_I|}{U_{RP_0} - U_{RP_1}}$ (Eq. 11)

where, P_C is the extra amount of money that a consumer has to pay for the improved version as compared with that of the base (regarding the mobile phone the compound attribute for the base version costs 90 euros and for the improved it costs 180 euros so P_C is 90 euros – see table 5), U_B the utility for the compound attribute of the base version, U_I the utility for the compound attribute of the improved version, U_{RP_0} the utility of the initial level of the reservation prices and a is the scale factor which represents the scaling differences of reservation prices. The value of the scale factor was determined to in order to satisfy the following design criteria:

- The profile which represents the base version should always receive the lowest total price
- The profile which represents the base version should always be cheaper than the one that represents the improved.
- The profiles of the base and improved versions should have substantial different prices

Moreover, every utility represents a x_{ijk} score which is an individual attribute level part-worth estimate as explained in Section 3.

For the mobile phone:

- For Equation 9, if the difference of the utility gained from the compound attribute level which represents the improved version minus the utility gained from the compound attribute level which represents the base version is positive this means that the particular respondent prefers the improved version over the base version.
- If the difference of the utility gained from the first level of reservation price minus the utility gained from second level of reservation price is positive this means that the particular respondent is willing to pay the extra amount of money which represents the cost of the improved version plus a maximum amount of 19.999 euros, according to his income, to acquire the improved version.
- Then the WTP for that respondent can be determined by the sum of the extra amount of money that a consumer has to pay for the improved version (as being defined by the compound attribute

cost) and a constant which is described by the ratio of the difference in the utilities of improved and base version (numerator) divided by the difference in the utilities of the first and second reservation price levels (denominator) multiplied by the scale factor which is determined according to design as explained before.

- The constant ensures that the solution of Equation 9 cannot exceed the range of 90 euros and 109.999 euros. The immediate next solution range is being defined by the amount of 110 and 129.999 euros which is the case of Equation 10.
- For Equation 10, if the utility gained from compound attribute level for the improved version minus the utility gained for the compound attribute level for the base version is between the utilities gained from the differences of the first reservation price level minus the second reservation price level and the first reservation price level minus the third reservation price level this means that the particular respondent is willing to pay the extra amount of money which represents the cost of the improved version plus a maximum amount of 39.999 euros.
- However, according Equation 11, if the difference of the utility gained from the compound attribute level which represents the improved version minus the utility gained from the compound attribute level which represents the base version is negative this means that the particular respondent prefers the base over the improved version.
- Then, if the difference of the utility gained from the first level of reservation price minus the utility gained from second level of reservation price is positive this means that the particular respondent is willing to pay the extra amount of money which represents the cost of the improved version minus a maximum amount of 19.999 euros, according to his income, to acquire the improved version (since the formulas represent how much a respondent is willing to pay for the improved version).
- Briefly, these formulas count how much money the respondent is willing to pay for an improved version over the base version. Equation 9 represents the reverse modelling of the situation in order to provide WTP values even in the case when the respondent is not willing to pay additional money from his income to buy the improved version. Particularly, in that case it is assumed that he would be willing to pay less than the cost of the compound attribute to acquire the improved version.
- As being described by Equations 9, 10 and 11 the solution range is determined by the scale factor, the compound attribute utilities for the improved and the base versions, the reservation prices and their utilities and the extra cost which is asked for the improved version. For the mobile phone the solution range lies within the values of 10 and 170 euros with intermediate steps 30, 50, 70, 90, 110, 130 and 150 euros, since the value of the scale factor for the is 20 euros.

For this procedure, the positive values which are higher than P_C plus the highest reservation price and lower than P_C minus the highest reservation price are meaningless and these scores, if any, should be collapsed within the maximum and minimum scores that satisfy the solution range. Moreover, the negative values should be collapsed to 0 and zero WTPs should remain as they are. Actually, the limits of the solution range for CBCA take place only for positive values since the formula calculates the extra amount of money that consumers are willing to pay for an improvement versus the status quo. However, WTP can be negative or zero meaning that consumers are not willing to pay.

The piecewise linear interpolation is the major reason for employing the second step of the procedure which was mentioned in the beginning of this Section, regarding the comparability of the results. In order to empirically compare the individual WTPs for each method the price range should be the same in each method. So, for the mobile phone the price range for positive WTPs was determined to be euros and 170 euros. So, the positive WTP values of the other two methods (CV and PCPM) should lie within the same range. So, the differences of the three methods on their distributions and means can be tested for a specific range of prices, otherwise the comparisons will be of no statistical importance.

The demand curves for all three methods were created in SPSS software, by with the Survival Tables tool. Survival functions of the form $q(p) = \Pr(p \leq WTP)$ were used. Here, $q(p)$ denotes the probability that a respondent's WTP is greater than a certain price level p .

5.3.2 Paired Comparison-Based Preference Measurement

Firstly, paired-comparison matrices were created for every sub-problem for every individual. According to the technical appendix of Scholz et al. (2010), the eigenvectors of each matrix were calculated representing the utilities of each attribute and attribute level. Then, the utilities of each attribute level belonging to a particular attribute were normalized according to the total number of levels in this attribute. Then, the utility of each attribute level was multiplied by the utility of the particular attribute. For the determination of WTP, a linear instead of piecewise linear interpolation was used, following a similar procedure with that of Kohli and Mahajan (1991). The procedure used to calculate individual WTPs can be explained as:

$$WTP = -\frac{(U_I - U_B)}{(U_P)}, \quad (\text{Eq. 12})$$

where according to the weighted additive model in equation 6,

U_I is the sum of the weighted utility scores of the attribute levels used to describe the improved version and U_B is the sum of the weighted utility scores of the attribute levels used to describe the base version. On the other hand U_P is the utility of price which was estimated by running linear regression with the weighted utilities of price levels as dependent variables and the price levels themselves (see Table 5) as independent.

5.3.3 Contingent Valuation: There is no need to further explain the WTP determination procedure for CV. It has already been clear that the price that a respondent accepts in the highest bid, it directly gives the WTP for that respondent.

6. Results

6.1 Mobile phone

Tables 8 and 9 represent the average utility scores derived for attribute levels in CBCA and PCPM for mobile phone. Since, CV is a direct method there are no utilities for the attribute levels included in the hypothetical scenario since WTP is directly stated from the respondents by the maximum bid they accept.

Table 8: Averages of the utility scores for attribute levels derived by CBCA for mobile phone

Black	White	Other	1.0GHz	1.3GHz	1.7GHz	5MP/16GB*	10MP/32GB**	12MP/32GB	10MP/64GB
0.880	0.772	-1.652	0.237	0.099	-0.336	0.572	3.976	1.177	-0.147

*Compound attribute level utility for the base version

** Compound attribute level utility for the improved version

12MP/64GB	RP_level 1	RP_level 2	RP_level 3	RP_level 4	RP_level 5	None
-5.579	0.515	-0.268	-0.31	-0.433	0.496	0.04

Table 9: Averages of the weighted utility scores for attribute levels derived by PCPM for mobile phone

Price	Mem	Cam	€350w*	€450w	€550w	€650w	5MPw	10MPw	12MPw
0.473	0.339	0.189	0.978	0.463	0.305	0.144	0.088	0.169	0.309

16GBw	32GBw	64GBw	Base	Improved	Price Coefficient
0.215	0.339	0.462	0.303**	0.508***	-0.003****

*w stands for weighted meaning that the utility of the attribute level (€350) is weighted by the utility of the whole attribute (Price)

** $U_B = 0.088(5MPw) + 0.215(16GBw) = 0.303$ *** $U_I = 0.169(10MP) + 0.339(32GBw) = 0.508$

**** Price coefficient after regressing price levels themselves on price level utilities

By looking the average utility scores of base and improved version in both Tables 8 and 9 it becomes quite obvious that respondents prefer the improved version over the base one in both methods. Regarding prices, PCPM has a negative price coefficient (utility) meaning that as the product becomes more expensive the respondents' preference will decrease. This can be verified also by the average weighted price levels in PCPM where the structure of preferences regarding price is presented as €350 > €450 > €550 > €650 (where > means that one price level is preferred over the other). However, the value of this coefficient is quite small meaning that the effect of price on preference about the stimulus is very small.

For CBCA, the situation about average reservation price levels is more trivial. The first reservation price level has the highest positive utility and this is logical since for mobile phone the first reservation price level is 0 (see Table 5). Also, for the rest three reservation price levels the utility scores are expectable since they indicate the decrease in utility gained as the price increases. However, the next highest positive utility score it is gained for the fifth reservation price level which is the most expensive meaning that for some respondents the increase in price increases also their satisfaction and thus their utility for the stimulus. So, there is probably a distinct cluster of respondents that relate price with quality, so they are willing to pay high prices to ensure the quality of the product.

Figure 6 presents the observed demand of WTP for the mobile phone improvement in each hypothetical method, which is measured as the amount of respondents whose WTP is greater than a given price level. The negative and zero WTP scores were collapsed to 0 for all the methods representing the amount of respondents who are not willing to pay for an improvement. Thus, WTP scores are classified as no positive (zero and negative values) and positive.

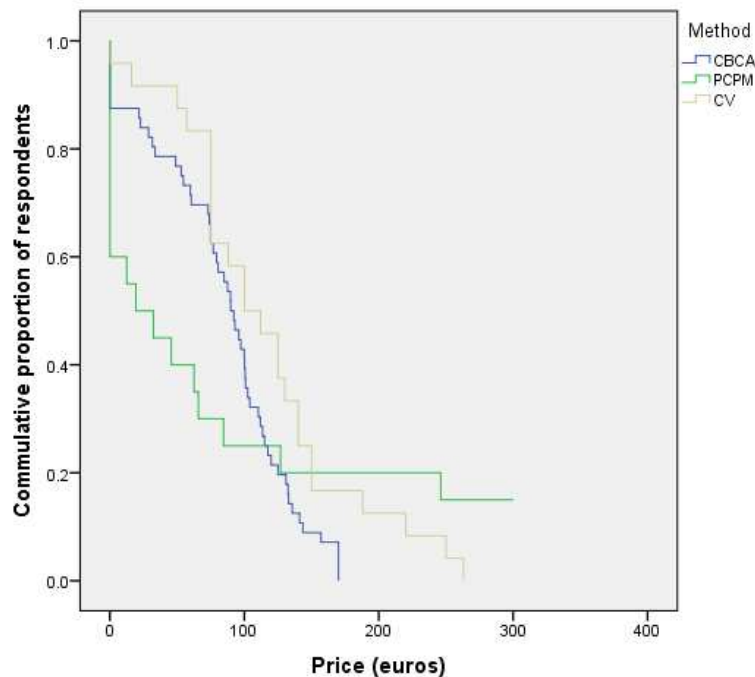


Figure6: Demand curves for mobile phone for each WTP method

The WTP determined via the means of Choice Based Conjoint Analysis is distributed quite smoothly and the scores are more differentiated as compared to the others, according to Wertenbroch and Skiera (2002) this suggests that the distribution of method has the best accuracy. Moreover, in the graph the limitation of the piecewise linear approach becomes obvious since CBCA cannot predict values smaller than 10 and greater than 170. Concerning the WTP for mobile, this does not seem to devalue the performance of the method. CV exhibits also a smooth distribution of WTP and it appears to be able to predict a larger range of scores as compared to that of CBCA. But, the curve has a fatter tail and the scores are not as differentiated as in CBCA which means less accuracy leading to the inference that CBCA captures more variation in responses whereas WTP in CV sticks in major price points which makes it less accurate.

The distribution of WTP for PCPM shows the least accuracy as compared with that of CV and CBCA. The distribution of PCPM's WTP makes two breaks which makes its interpretation a challenging task. Initially, 40% of respondents appeared to be unwilling to pay for the improvement with WTP scores equal to 0. For the other 40% of respondents who are willing to pay up to (approximately) 90 euros the WTP scores seem to be as smooth and differentiated as in CV. Additionally, regarding positive WTP up to 90 euros, PCPM appears to provide the most conservative WTP scores, as compared with CV and CBCA, because it has the steepest demand curve up to the second break point. But, for rest of 20% of respondents the tail of the demand curve becomes extremely fat as compared with the others, meaning that PCPM cannot predict well WTPs which are higher than 90 euros. So, there are respondents in PCPM that appeared to be willing to pay almost an infinite amount of money.

As mentioned the in the previous section the price range of WTPs for all three methods should be the same, otherwise any statistical comparison would be meaningless. Figure 13 presents the histograms of WTP scores per method after collapsing all the positive values of CV and PCPM within the price range of CBCA and the negative ones to 0. The histograms are alternative representations regarding the distribution of the observed demand. In

Figure 7 it becomes clear that CBCA can capture the most differentiated responses as compared with CV and PCPM, even within its constrained price range.

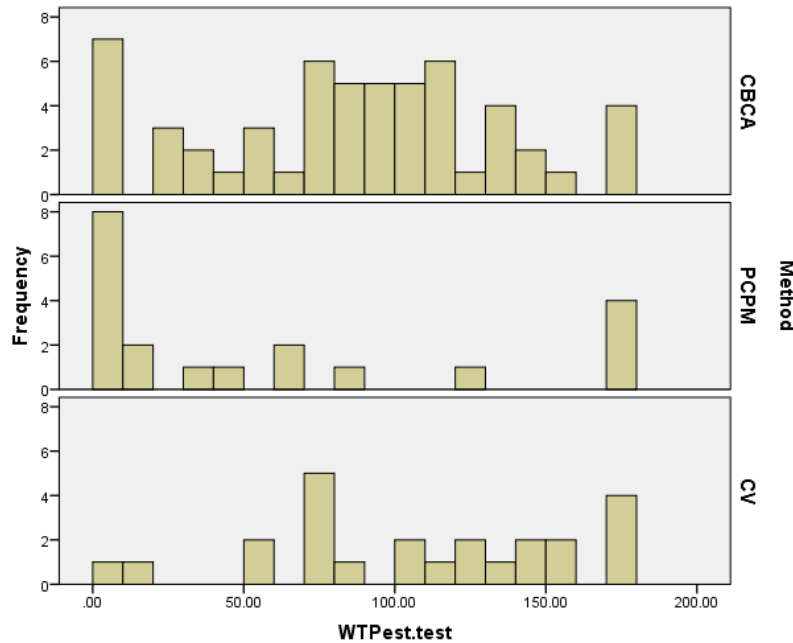


Figure7: Histograms of WTP scores for mobile phone after collapsing the maximum and minimum scores of CV and PCPM within the solution range of CBCA

CV captures a good amount of variation in responses, but it seems to provide a more optimistic prediction since the amount of no positive WTP scores is very small, this can be seen also in Table 10 which describes the portions of positive and no positive WTPs for each method. On the contrary, PCPM has the highest amount of no positive WTPs.

Table10: Percentages of negative and positive WTPs for each method for mobile phone

	CBCA	PCPM	CV
No positive WTP %	12.5	40	4.2
Positive WTP %	87.5	60	95.8
Total N	56	20	24

Also, regarding PCPM, there is a solid tendency to extreme values. This tendency can be verified also in Table 11 which presents the mean WTP and standard deviation of WTP for each method. PCPM has the highest standard deviation which is a result of extreme WTP scores. According to the histograms in Figure 7, one can also see an inclination of CV towards high positive WTP scores. CBCA seems to be the most realistically distributed with most of the WTP scores concentrated around the middle point of its price range, meaning that most of the respondents are willing to pay some amount of money to gain an improved version of a mobile phone.

Table11: Mean and Standard Deviation for each WTP method for mobile phone

WTP Method	Mean	SD
CBCA	84.88	48.83
PCPM	56.47	67.65
CV	105.75	48.44

Regarding mean WTP (Table11), PCPM gives the lowest mean as a result of the large amount of no positive WTP scores. On the other hand CV gives the highest mean WTP as a result of the small amount of no positive WTP scores. Judging from an overall consideration of the results, CBCA seems to be able to provide the most realistic mean regarding WTP for an improvement of a mobile phone. Because the data did not allow for testing the mean WTP differences among the three methods, only the WTP distributions were empirically compared. Particularly, the assumption of normality was not satisfied for all the methods making the application of parametric tests, like t-test or analysis of variance (ANOVA), impossible (see Appendix C).

To empirically compare WTP distributions of the three methods, nonparametric tests were applied. Initially, the Kruskal–Wallis (KW) test was performed in order to compare all distributions at once. Then, Mann-Whitney (MW) test, two-sided Kolmogorov-Smirnov (KS) test, and Siegel and Castellan (S&C) procedure were applied in order to compare pairwise every method with each other. The results of each test are presented in table 12.

Table12: Test results comparing the WTP distributions of CBCA, PCPM, and CV

	KW	MW	KS	S&C
CBCA_PCPM_CV	0.016	N.A	N.A	N.A
CBCA-PCPM	N.A*	N.S**	N.S	N.S
PCPM-CV	N.A	.01	.004	N.S
CV-CBCA	N.A	N.S	N.S	N.S

*N.A: not applicable, meaning that the specific test cannot be used for the specific comparison, **N.S: not significant at 0.0167 level of significance after applying Bonferroni correction

Regarding KW test, the WTP distributions of the three methods for the mobile phone improvement were significantly different, $H(2) = 8.27$, $p < .05$. For MW and KW tests a Bonferroni correction was applied so all effects are reported at a .0167 level of significance. The application of Bonferroni correction is to adjust the level of significance according to the total number of comparisons. Regarding MW test, it appeared that WTP distributions were not significantly different among CBCA and PCPM ($U = 384$, $r = -.24$). Also, the WTP distributions were not significantly different among CBCA and CV ($U = 519.5$, $r = -.18$). However, WTP distribution of CV was significantly higher as compared with that of PCPM ($U = 131$, $r = -.4$) with a moderate effect size. Test statistic U refers to a nonparametric test of the null hypothesis that two samples come from the same population against an alternative hypothesis that the two samples come from different populations. Particularly the U statistics it is used to indicate if a one population tends to have higher values than the other. Moreover, r, it represents the effect size.

Regarding KS test, the results are exactly the same. Further, S&C procedure was applied, which is more stringent from MW and KS because except the correction for the total number of comparisons it considers also the total

number of respondents examined under the two methods which are compared. For this procedure, there were no significant differences among the WTP distribution regarding each pair of the methods.

The critical decision to be taken in order to draw some conclusions according to the differences between WTP distributions, it has to do with what test one should consult. Regarding the majority of the non-parametric tests about WTP distributions for the mobile phone improvement, it can be concluded that PCPM distribution is significantly different as compared with that CV. Nevertheless, there were no significant differences in WTP distribution regarding CBCA and CV, and CBCA and PCPM.

6.2 PC mouse

Tables 13 and 14 represent the average utility scores derived for attribute levels in CBCA and PCPM for mobile phone.

Table 13: Averages of the utility scores for attribute levels derived by CBCA for PC mouse

Erg. Y	Erg. N	Guar. Y	Guar. N	AirButtons NN*	AirButtons YN	AirButtons_NY	AirButtons_YY**	RP level 1	RP level 2
0.025	-0.025	-0.131	0.131	0.334	-0.066	0.207	-0.476	0.0428	0.0428

*Compound attribute level utility for the base version

** Compound attribute level utility for the improved version

RP level 3	RP level 4	None
0.0428	-0.128	-1.60498

Table 14: Averages of the utility scores for attribute levels derived by PCPM for PC mouse

Price	Buttons	Air Motion	w*€15	w€30	w€45	w€60	wExtraButtons Y	wExtraButtons N*	wAirMotion Y
0.912	0.755	1.245	1.245	0.734	0.319	0.172	0.156	0.147	0.166

wAirMotion N	Basic	Improved	PriceCoefficient
0.294	0.441*	0.322**	0.0003***

*w stands for weighted meaning that the utility of the attribute level (€15) is weighted by the utility of the whole attribute (Price)

** $U_B = 0.147 (wExtraButtonsN) + 0.294 (wAirMotionN) = 0.441$

*** $U_I = 0.156 (wExtraButtonsY) + 0.166 (wAirMotionY) = 0.322$

**** Price coefficient after regressing price levels themselves on price level utilities

In contrast with the results of the mobile phone, the improved version it is not preferred over the base version regarding the respondent in the PC mouse study. For PCPM the utilities of average weighted price levels and price coefficient seem contradictory. Regarding the coefficient of price the respondents' preferences about the stimulus appeared to increase as the price increases, since there is a positive sign. Nevertheless, the value of price coefficient is too small which suggests interpreting respondents' attitude towards price to be neutral in PCPM. On the other

hand, the average weighted price level utilities in PCPM exhibit the same preference structure with that of mobile phone, meaning that cheaper price is preferred. By combining the information from the regression coefficient of price, the average weighted price level utilities and the average utilities of the base and improved version it can be inferred that price would not prevent the respondents from buying the product. However, it is the improvement itself that does not appear to add value for these respondents. So, respondents are not willing to pay for the improved version not because of price but due to the fact that they do not see their benefits increased.

In CBCA respondents seem to prefer higher prices but up to a certain extent. These respondents seem to perceive increase in price to be related with higher quality, as in mobile phone, however their preferences decrease when they are asked to pay the highest price (see reservation price level utilities – Table 13) This is not weird since PC mice are generally low cost products so respondents who would like to acquire a version with extra features would be willing to pay for it without caring much about the extra cost. So, concerning both PCPM and CBCA utilities derived for PC mouse, the attribute levels themselves appear to be the determinants for respondents' preferences and not the price.

Figure 8 presents the observed demand of WTP for the mobile phone improvement in each hypothetical method. On the contrary with the results of the mobile phone, the demand curve of CV appears to outperform that of CBCA. The demand curve of CV is the smoothest as compared with the others. CBCA predicted demand still captures the most differentiated responses as compared with that of CV and PCPM (Fig. 9), although the limitation of the method regarding its constraint price range, due to piecewise linear interpolation, it becomes quite harmful for its demand curve.

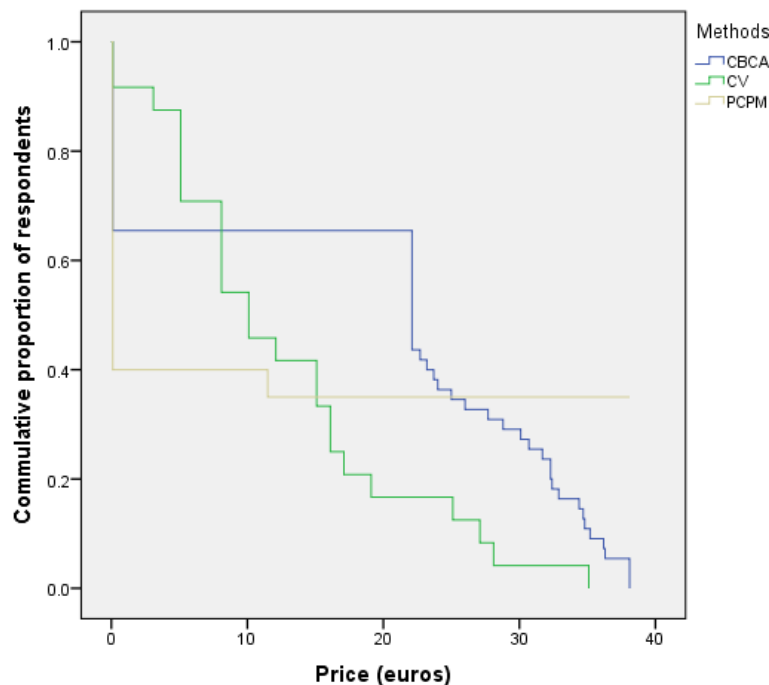


Figure8: Demand curves for PC mouse for each WTP method

CBCA cannot predict values smaller than 22 and greater than 38 euros (Fig. 8). The maximum value of this range does not seem to create a problem in the prediction of high positive WTP scores, although CBCA has a fatter tale as compared with that of CV (Fig. 8 and Fig. 9). The minimum value of the range though, it is quite restrictive since CBCA cannot predict the amount of the respondents with low positive (less than 22 euros) WTP scores.

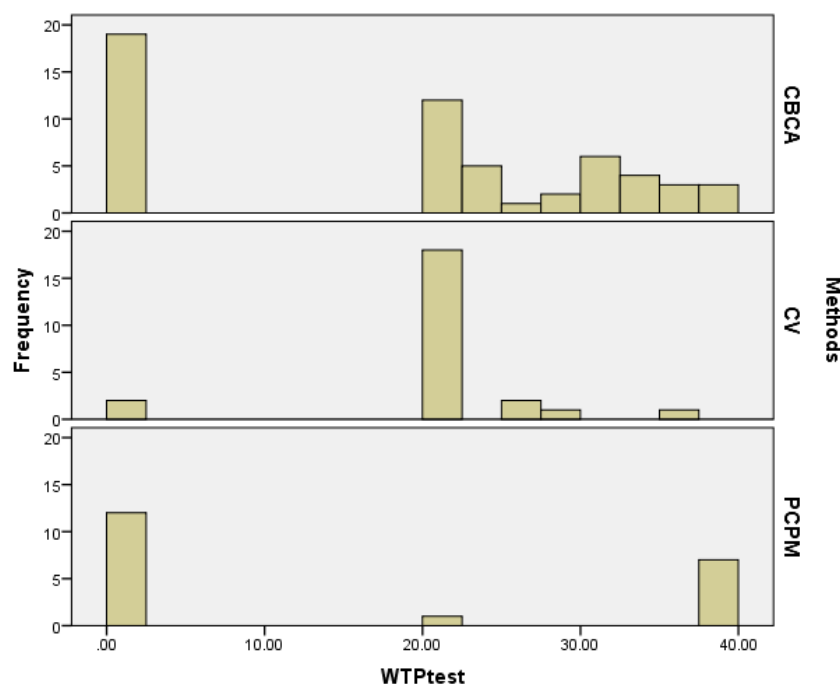


Figure 9: Histograms of WTP scores for PC mouse after collapsing the maximum and minimum scores of CV and PCPM within the solution range of CBCA

CBCA shows a large amount of respondents that they are not willing pay for a PC mouse improvement (Table14) as compared with that of mobile phone. This is expectable since, in line with the results derived from HB logit, the CL logit coefficient for the compound attribute level representing the improved version was not significant. This means that CBCA for the PC mouse cannot predict well WTP, since the improvement did not appear to be important for the specific sample (for further discussion regarding the results of the CL model see Appendix D). The amount of no positive WTPs regarding CV is twice as big as compared with that of the mobile phone, but it is still very small.

In PCPM there is a dominant tendency to extreme scores (Fig 9). Also, PCPM provides extremely conservative WTP scores with 60% of the respondents showing no positive WTP. This is not strange since the method just indicates that most of the respondents are not willing to pay for a PC mouse improvement. However, PCPM has the least informative distribution of WTP as compared with those of CV and PCPM. Also, PCPM has the largest standard deviation followed from CBCA (Table15). CV appears to have the smallest standard deviation.

Table14: Percentages of negative and positive WTPs for each method for PC mouse.

	CBCA	PCPM	CV
No positive WTP %	34.5	60	8.3
Positive WTP %	65.5	40	91.7
Total N	55	20	24

Therefore, CV provides the most realistic distribution of demand as compared with PCPM and CBCA because the method has the smoothest demand curve (Figures and 9) and thus the most accurate. Moreover, CV has the smallest standard deviation (Table 15).

Concerning mean WTPs (Table 15) the results are the same with those of mobile phone with CV having the highest mean and PCPM the lowest. On the contrary with the results of mobile phone, PCPM provides the most realistic mean since respondents did not appear to prefer the improved version over the base version regarding the utilities derived from CBCA and PCPM. This information leads to the inference that in general respondents are not willing to pay extra money to acquire the improved version of the PC mouse. So, the conservative mean of PCPM seems to represent reality better than the mean of the other two. Thus, regarding the determination of mean WTP PCPM is the best method. Again the data did not allow for comparing mean WTP across the methods.

Table15: Mean and Standard Deviation for each WTP method for PC mouse

WTP Method	Mean	SD
CBCA	18.4	14.32
PCPM	14.4	18.42
CV	21.29	7.21

Table 16 provides the results of the non-parametric test conducted for the WTP distributions of the PC mouse. The same tests and correction procedures for the mobile phone were applied also for the PC mouse.

Table16: Test results comparing the WTP distributions of CBCA, PCPM, and CV for PC mouse

	KW	MW*	KS*	S&C
CBCA_PCPM_CV	N.S*	N.A	N.A	N.A
CBCA-PCPM	N.A*	N.S.	N.S.	N.S.
PCPM-CV	N.A	N.S.	.006	N.S.
CV-CBCA	N.A	N.S.	N.S.	N.S.

*N.A: not applicable, meaning that the specific test cannot be used for the specific comparison, **N.S: not significant at 0.0167 level of significance after applying Bonferroni correction

Regarding KW, MW tests and S&C procedure the WTP distributions of the three methods for the mobile phone improvement were not significantly different. Regarding KS test, it appeared that WTP scores of PCPM were significantly lower from those of CV ($p < .0167$).

Regarding the majority of the non-parametric tests about WTP scores for the PC mouse improvement, it can be concluded that there were no significant differences in WTP distributions regarding the three methods. However, KS

is the most commonly used test for comparing WTP distributions (Millera et al., 2011; Silva et al., 2007) and it is more powerful than MW for samples smaller than twenty five respondents. Although KW, MW and S&C do not show any significant difference regarding the distributions of PCPM and CV for PC mouse, the significant difference found in KS test regarding these distributions is in line with the results of mobile phone. Therefore, for both products the WTP distributions of CV and PCPM are significantly different in terms of WTP distribution of PCPM being lower than that of CV. There is no evidence regarding other significant differences in distributions for the rest of the empirical comparisons among the hypothetical WTP methods.

7. Discussion

This study has several implications regarding the hypothetical WTP methods. First, the results indicate that CV appears to be the most consistent method regarding the information that can be extracted from its distribution. Although CV appeared to be outperformed from CBCA in the mobile phone scenario, the distribution of the method appeared to be quite informative for both products. This means that the method can be used to elicit WTP for different product categories.

Second, the piecewise linear interpolation appeared to be very harmful for CBCA in the calculation of WTP for the PC mouse. This could be caused for two reasons. The one reason is that piecewise linear interpolation was employed in order to uncover the effects of different reservation price levels. According to Jedidi and Zhang (2002), reservation prices imply that consumers consider their income when they are thinking of buying or not buying a product. So, the inclusion of different reservation prices increases the possibility to capture different income levels. However, consumers might not consider their income for a product like a PC mouse. So the inclusion of reservation prices for determining WTP for a cheap product did not appear to be useful. The other reason is the very limited solution range defined by piecewise linear interpolation for PC mouse improvement leading to inferior prediction of low positive WTP values. In other words, the use of piecewise linear interpolation and reservation prices are not recommended in CBCA for the determination of WTP, regarding cheap products. Nevertheless, this is not the case for expensive products like a mobile phone, where CBCA appeared to show the best accuracy.

Third, the demand curves derived by PCPM did not appear to be very informative as compared with CV and CBCA, concerning both products. The major shortcoming of the method is the solid tendency to extreme scores for mobile phone which becomes dominant for PC mouse. According to Scholz et al. (2010), one of the major characteristics of PCPM is that leads to very distinctive attribute importances as compared with other methods. As a consequence PCPM results to large utility differences between the most and the least preferred levels. According to the results of this study, PCPM can be appropriate if the researcher is interested mostly in the mean WTP and the relative part-worths (utilities) of product attribute and price levels, but not for creating demand curves. By considering Eq. 12, one can infer that big differences in utilities of attribute levels will result to extreme WTP scores.

However, for the PC mouse, the large differences in utilities might not be the only reason for the dominance of extreme scores. The price range (35euros) used in PCPM for determining WTP for the PC mouse improvement it was very narrow as compared with the price range (300euros) used for determining WTP for a mobile improvement (see Table 5, price levels in PCPM). According to Parducci (1974) a narrow price range will decrease the average stimulus responses and increase the extreme responses. Also, PCPM was introduced in order to elicit preferences for complex products with large number of attributes and attribute levels. In this study the number of attributes and their levels was considerably small. This might be also a serious reason for the inferior performance of the method regarding its demand curves as compared with those of CV and CBCA. Unfortunately, there is no evidence in the literature of what would have happened if a different price range and a larger number of attributes and attribute

levels were employed. However, it is expected that a wider range of price and a larger number of attribute and attribute levels would have led in more informative distributions.

Fourth, the results of the empirical comparison concerning CBCA and CV differ with those reported by Miller et al. (2011) and Grunert et al. (2009). In the study of Miller et al. (2011) the demand distribution and mean WTP which were determined under CBCA were found to be significantly higher than those determined under those in CV. However, in the study of Miller et al. (2011) WTP was determined as the amount of money that a respondent is willing to pay for buying a product versus not buying a product which was not the case in this study since WTP was determined as the extra amount of money a respondent is willing for buying an improved version over a base version. Also, in Miller et al. (2011) CBCA appeared to have significantly higher mean and distribution than those in CV. In this thesis the mean and distribution of WTP appeared to be higher, although not significantly, for CV than CBCA. This difference in the results can be due to reservation prices which mean that their inclusion in CBCA can really make a difference in the results by considering also the information of consumers' idiosyncratic nature and different income levels. So, by including reservation prices in CBCA makes the WTP estimation more conservative in an effort to make it also more realistic. Moreover, in the study of Miller et al. (2011) CV was modelled as a logit form and utilities were estimated also for this method which again was not the case in this thesis.

The study of Grunert et al. (2009) also exhibited that the determined WTP for the improvement differ significantly regarding CBCA and CV. Grunert et al. (2009) they treated WTP in a similar way with this thesis, which was determined as the additional amount of money that a respondent is willing to pay for an improved version compared to the base version. However, in the study of Grunert et al. (2009) game money (versus real money) were used for the hypothetical elicitation of WTP (versus actual elicitation of WTP) and then analysis of variance was employed to test the differences of CBCA and CV on WTP after controlling for the effect of this experimental factor considering also other covariates which again was not the case concerning the statistical modelling of specific method effects in this thesis. Nevertheless, in the study of Grunert et al. (2009) the mean WTP for CBCA found to be lower than that of CV which is the same result that was found also in this thesis. It seems that when the income of respondents is being modelled within a CBCA design for the determination of WTP, either with reservation prices or with game money, then CBCA becomes more conservative than CV regarding their mean WTP scores.

Another reason for not finding significant differences between the distributions of CV and CBCA could be the compound attribute that was created for attaining a single utility score for the attribute levels under examination. Furthermore, the constrained price range that was employed in all three methods for conducting the statistical analysis can be another reason of no significant differences between CBCA and CV. Additionally, the distribution of WTP appeared to be significantly lower in PCPM as compared with that in CV. This quite logical since for both products PCPM provided the lowest mean and CV the highest mean.

This study also has several limitations. For CBCA, five choice sets in the mobile phone scenario and four choice sets in the PC mouse scenario were employed. Although this appeared to be sufficient for the specific study, using four or five choice sets is on the lower side of practice. The samples of CV and PCPM for both products were comprised by the minimum amount of people needed to conduct a statistical analysis. The WTP scores determined by each method were not compared with real data, as in Miller et al. (2011) in order to make inferences about their external validity. The two product categories used in this thesis are not the only ones. Hence, other product categories can be examined for their effects on the methods used for determining WTP, such as the fast-moving consumer goods and luxury goods. Also, the specific product formation used in PCPM prevented the smooth distribution of responses.

8. Conclusion

Through a large scale experimental design and field test this thesis compared directly measured WTP (CV), indirectly measured WTP determined by de-compositional manner (CBCA), and indirectly measured WTP determined by compositional manner (PCPM). This 3 in 1 study enabled a comprehensive examination on the differences of the three hypothetical methods in the way they elicit preferences to determine WTP.

Concerning the demand curves, in line with the suggestions from Miller et al. (2011), this study verifies the argument that direct approaches (CV) outperform indirect ones (CBCA) in the determination of WTP for cheap products, whereas indirect approaches (CBCA) outperform direct ones (CV) in the determination of WTP for expensive products. Therefore, it has been evident that the product in question does affect the results of each method regarding their demand curves. These conclusions are majorly derived by the smoothness and differentiated responses of the demand distributions created by the individual WTP scores of each hypothetical method.

The piecewise linear interpolation and the inclusion of reservation prices in CBCA profiles proved to be harmful for determining WTP for the PC mouse improvement. PCPM appeared to exhibit the least informative distribution for both products. The appearance of large utility differences in PCPM was the major reason of the tendency to extreme scores. Overall, CV demand curve, for both products, it appeared to be the most consistent as compared with the other two.

In contrast with the current literature, the WTP distributions of CBCA and CV were not proved to be significantly different. However, the technical modulations that were conducted in order to create a common frame for comparing these methods in a statistically meaningful way, they might cause the convergence of WTP among methods instead of their difference. Similarly to Silva et al. (2007), this study suggests that the selection of a specific WTP method should be based among the aim and the nature of research and that market researchers should perceive these methods as complementary rather than substitutes.

Regarding mean WTP, although there cannot be any statistical claim due to the nature of the data, this study shows that in general indirect approaches outperform the direct ones for both cheap and expensive products. For mobile phone CBCA appeared to have the most realistic mean and for PC mouse PCPM appeared to have the most realistic mean. Also, the order of the mean scores does not change according to the product, meaning that PCPM has always the lowest mean and CV the highest. So, the product in question does not affect the results of each method regarding their mean scores.

Concerning the current literature, there is no clear indication regarding the selection of the mean versus the demand curve in order to judge the performance of an approach in determining WTP. According to Miller et al. (2011), mean is an important measure of WTP. But, even an accurate measurement of mean WTP may not be sufficient since the demand curve may imply a very different pricing structure, like in the application of PCPM in determining WTP for the PC mouse. Therefore, researchers must also consider the entire WTP distribution in assessing the performance of an approach and not just the mean.

The results of this thesis lead to several recommendations for further research. Initially, for CV WTP was determined by the maximum bid that a respondent accepts. So, the higher price that a respondent accepts was treated as the determined WTP for that respondent. Studies like those of Miller et al. (2011) and Welsh and Poe (1998), among many others, model CV by using a logit form like that in CBCA. So, it is recommended to empirically compare the WTP determined from CV via the means of discrete choice modelling with the WTP determined from other methods. The low number of attributes and attribute levels used in PCPM it is the most plausible reason for its unsatisfactory performance concerning its demand curves, especially for the PC mouse. Therefore, more complex

stimulus should be used in order to draw conclusions regarding the performance of PCPM as compared to other WTP methods. Also, it would be interesting to use piecewise linear interpolation for calculating WTP in PCPM. Moreover, an experimental condition like that of game money should be used to test any design effects of the hypothetical methods on WTP.

Moreover, the individual utility scores and determined WTP via PCPM and CBCA for both products indicated that there are respondents who:

- prefer the improved version and low prices, and therefore they have a positive WTP (expected)
- prefer the base version and low prices, and therefore they have a negative WTP (expected)
- prefer the improved version and high prices, and therefore they have a negative WTP (different perception of price-unexpected)
- prefer the base version and high prices, and therefore they have a positive WTP (extremely high WTP for the base version-unexpected)

The value of WTP is a function (Equations 9, 10, 11 and 12) of the signs and values of utilities derived for price and compound attribute levels. This distinction of respondents uncovers two categories of respondents with expected and two categories with unexpected WTP scores. This is an indication of how to distinguish respondents in this study which is a result of the harsh scrutinizing of data, but the low amount of respondents did not allow for further statistical analysis. It is suggested though, that cluster analysis should be employed in a similar study to uncover different segments of respondents by combining utilities with demographics. Then, these potential clusters would be enriched with additional information and the interpretation regarding their WTP would be more beneficial.

Finally, there is a significant gap in the literature regarding a sufficient theoretical underpinning for explaining any method effects on WTP. Also, there is a major research priority to shed light on the cognitive processes of consumers when they state or reveal their preferences (Grunert et al, 2009). The study of Grunert et al. (2009) drew insights from price information processing research to derive three constructs that characterize the cognitive mechanism of consumers in order to measure the effects of these constructs on WTP. Particularly, the first construct is the price involvement which represents the eagerness of consumers to search for and process price information. The second construct is the reference price which represents previously stored information on the existing prices of the product in question or similar products. The third construct refers to choice heuristics that which represent the way that consumers combine price with other information to make a purchase decision. So, the introduction of additional method-specific constructs, and the way all these constructs affect WTP should be thoroughly examined by proposing and testing a conceptual framework. This theoretical framework would be a significant contribution on theory development regarding any method-specific and person-specific effects on WTP.

The answers to these questions should help on postulating which methods generate the most accurate WTP for a broader range of circumstances than those examined under the current literature.

Appendix

Appendix A

Instructions given for the improvement of PC mouse:

The following pair comparisons might contain characteristics that are difficult for you to understand. For your convenience the following terms are explained:

- **Air Motion:** For a PC mouse, air motion means that a person with his natural hand movements can handle applications like spreadsheets, word documents, and presentations. This mobile mouse is wireless and enables intuitive on- and off-the-desktop computing.
- **Extra Buttons:** Extra refers to additional keys except the right, the left, and the scroll wheel. These keys can be used for performing additional PC functions from the mouse without using the keyboard, like volume setting, internet surfing, or gaming among many others.

See also the two pictures in order to better realize the differences between an ordinary PC mouse and a PC mouse with Air Motion and Extra Buttons.



ordinary mouse



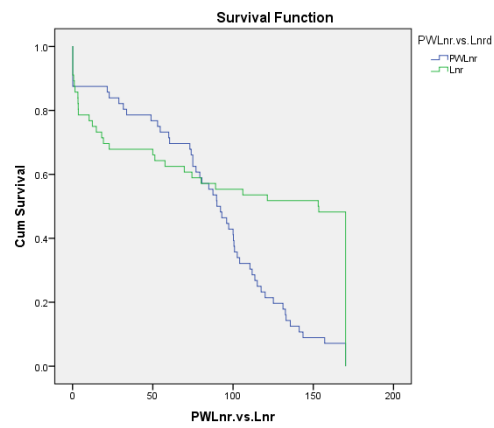
mouse with air motion and extra buttons

Appendix B

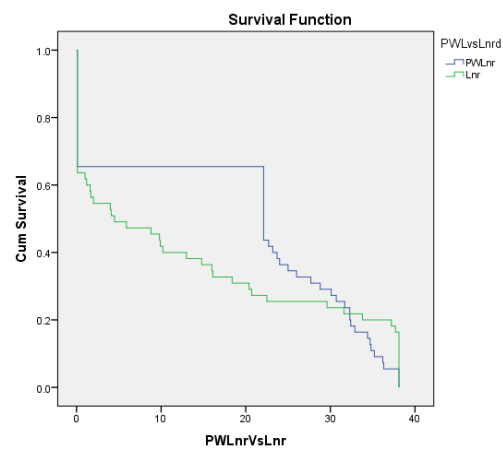
Representation of survival graphs regarding the demand curves created by linear and piecewise linear approach:

In both cases it is obvious that linear approach in CBCA results in fat tails.

Mobile phone



Pc mouse



Appendix C

Kolmogorov-Smirnov for checking normality:

When the Kolmogorov-Smirnov test is not significant then normality can be assumed.

Mobile phone	KS
CBCA	.200
PCPM	.023
CV	.200

PC mouse	KS
CBCA	.000
PCPM	.000
CV	.000

Appendix D

Results from the Conditional Logit:

Regarding the results from CL on can check that almost every attribute level is significant for Mobile, and no attribute level is significant for PC mouse (the results are reported in .05 levels of significance).

Mobile phone

Attribute levels	Significance
dwhite	0.0036
dblack	0.08
d13GHz	0.04
d17GHz	0.014
d1032	0.000
d1064	0.0015
d1232	0.00049
d1264	0.0089
PRICE	0.000

PC mouse

Attribute levels	Significance
dergonomic	0.92
d2years	0.68
dny	0.92
dyn	0.8
dyy	0.9
PRICE	0.75

References

- 1) Aizaki, H. (2012). Basic functions for supporting an implementation of choice experiments in R. *Journal of Statistical Software*, 50(2), 1-24.
- 2) Adamowicz, W., Boxall, P., Williams, M., & Louviere, J. (1998). Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American Journal of Agricultural Economics*, 80(1), 64-75.
- 3) Aizaki, H., & Nishimura, K. (2008). Design and analysis of choice experiments using R: a brief introduction. *Agricultural Information Research*, 17(2), 86-94.
- 4) Alberini, A., Kanninen, B., & Carson, R. T. (1997). Modeling response incentive effects in dichotomous choice contingent valuation data. *Land Economics*, 73(3), 309-324.
- 5) Alfnes, F. (2004). Stated preferences for imported and hormone-treated beef: application of a mixed logit model. *European Review of Agricultural Economics*, 31(1), 19-37.
- 6) Anderson, J. C., Jain, D., & Chintagunta, P. K. (1993). Understanding customer value in business markets: Methods of customer value assessment. *Journal of Business-to-Business Marketing*, 1(1), 3-30.
- 7) Arora, N., & Huber, J. (2001). Improving parameter estimates and model prediction by aggregate customization in choice experiments. *Journal of Consumer Research*, 28(2), 273-283.
- 8) Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., & Schuman, H. (1993). Report of the NOAA Panel on Contingent Valuation. 4601-4614
- 9) Balistreri, E., McClelland, G., Poe, G., & Schulze, W. (2001). Can hypothetical questions reveal true values? A laboratory comparison of dichotomous choice and open-ended contingent values with auction values. *Environmental and Resource Economics*, 18(3), 275-292.
- 10) Backhaus, K., Wilken, R., Voeth, M., & Sichtmann, C. (2005). An empirical comparison of methods to measure willingness to pay by examining the hypothetical bias. *International Journal of Market Research*, 47(5), 543-562.
- 11) Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3), 226-232.
- 12) Bunch, D. S., Louviere, J. J., & Anderson, D. (1996). A comparison of experimental design strategies for multinomial logit models: The case of generic attributes. *University of California Davis Graduate School of Management Working Paper*, 11-96.
- 13) Boccaletti, S., & Nardella, M. (2000). Consumer willingness to pay for pesticide-free fresh fruit and vegetables in Italy. *The International Food and Agribusiness Management Review*, 3(3), 297-310.
- 14) Cameron, T. A., & James, M. D. (1987). Estimating willingness to pay from survey data: an alternative pre-test-market evaluation procedure. *Journal of Marketing Research*, 69(2), 389-395.
- 15) Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research*, 44(2), 214-223.
- 16) Duffield, J. W., & Patterson, D. A. (1991). Inference and optimal design for a welfare measure in dichotomous choice contingent valuation. *Land Economics*, 67(2), 225-239.
- 17) Elrod, T., Louviere, J. J., & Davey, K. S. (1992). An empirical comparison of ratings-based and choice-based conjoint models. *Journal of Marketing Research*, 29(3), 368-377.
- 18) Forman, E. H., & Gass, S. I. (2001). The analytic hierarchy process-an exposition. *Operations Research*, 49(4), 469-486.
- 19) Green, P. E. (1984). Hybrid models for conjoint analysis: An expository review. *Journal of Marketing Research*, 21(2), 155-169.
- 20) Green, P. E., Goldberg, S. M., & Montemayor, M. (1981). A hybrid utility estimation model for conjoint analysis. *The Journal of Marketing*, 45(1), 33-41.
- 21) Green, P. E., Krieger, A. M., & Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, 31(3_supplement), S56-S73.
- 22) Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8(3), 355-363.

- 23) Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, 5(2), 103-123.
- 24) Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. *The Journal of Marketing*, 54(4), 3-19.
- 25) Grunert, K. G., Juhl, H. J., Esbjerg, L., Jensen, B. B., Bech-Larsen, T., Brunsø, K., & Madsen, C. Ø. (2009). Comparing methods for measuring consumer willingness to pay for a basic and an improved ready made soup product. *Food Quality and Preference*, 20(8), 607-619.
- 26) Huber, J., Orme, B. K., & Miller, R. (2007). Conjoint measurement. Methods and applications. *New York Springer*, 347-352
- 27) Hauser, J. R., & Shugan, S. M. (1980). Intensity measures of consumer preference. *Operations Research*, 28(2), 278-320.
- 28) Hartvigsen, D. (2005). Representing the strengths and directions of pairwise comparisons. *European Journal of Operational Research*, 163(2), 357-369.
- 29) Helm, R., Manthey, L., Scholl, A., & Steiner, M. (2003). Empirical evaluation of preference elicitation techniques from marketing and decision analysis (No. 02/2003). Friedrich-Schiller-Universität Jena, Wirtschaftswissenschaftliche Fakultät.
- 30) Helm, R., Scholl, A., Manthey, L., & Steiner, M. (2004a). Measuring customer preferences in new product development: comparing compositional and decompositional methods. *International Journal of Product Development*, 1(1), 12-29.
- 31) Helm, R., Steiner, M., Scholl, A., & Manthey, L. (2004b). A comparative empirical study on common methods for measuring preferences. *International Journal of Management and Decision Making*, 9(3), 242-265.
- 32) Hoffman, E., Menkhaus, D. J., Chakravarti, D., Field, R. A., & Whipple, G. D. (1993). Using laboratory experimental auctions in marketing research: A case study of new packaging for fresh beef. *Marketing Science*, 12(3), 318-338.
- 33) Hurley, W. J. (2001). The analytic hierarchy process: A note on an approach to sensitivity which preserves rank order. *Computers & Operations Research*, 28(2), 185-188.
- 34) Jedidi, K., & Zhang, Z. J. (2002). Augmenting conjoint analysis to estimate consumer reservation price. *Management Science*, 48(10), 1350-1368.
- 35) Johnson, R. M. (1987). Adaptive conjoint analysis. In *Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing* (No. 1, pp. 253-265). Sawtooth Software.
- 36) Russell, C. S., Krarup, S., & Clark, C. D. (2005). Environment, information and consumer behaviour: an introduction. 1-29.
- 37) Kohli, R., & Mahajan, V. (1991). A reservation-price model for optimal pricing of multiattribute products in conjoint analysis. *Journal of Marketing Research*, 28(3), 347-354.
- 38) Louviere, J. J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of Marketing Research*, 20(4), 350-367.
- 39) Louviere, J. J. (1988). Conjoint analysis modelling of stated preferences: a review of theory, methods, recent developments and external validity. *Journal of Transport Economics and Policy*, 22(1), 93-119.
- 40) Louviere, J. J., & Hensher, D. A. (1983). Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event. *Journal of Consumer Research*, 10(3), 348-361.
- 41) Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). Stated choice methods: analysis and applications. *Cambridge University Press*.
- 42) Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of mathematical psychology*, 1(1), 1-27.
- 43) Madansky, A. (1980). On conjoint analysis and quantal choice models. *Journal of Business*, 53(3), S37-S44.

- 44) Meißner, M., & Decker, R. (2009). An empirical comparison of CBC and AHP for measuring consumer preferences. In *International Symposium of Analytical Hierarchy Process*.
- 45) Meißner, M., Decker, R., & Scholz, S. W. (2010). An Adaptive Algorithm for Pairwise Comparison-based Preference Measurement. *Journal of Multi-Criteria Decision Analysis*, 17(5-6), 167-177.
- 46) Miller, K. M., Hofstetter, R., Krohmer, H., & Zhang, Z. J. (2011). How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research*, 48(1), 172-184.
- 47) Mitchell, R. C., & Carson, R. T. (1989). Using surveys to value public goods: the contingent valuation method. *Resources for the Future*.
- 48) Mulye, R. (1998). An empirical comparison of three variants of the AHP and two variants of conjoint analysis. *Journal of Behavioral Decision Making*, 11(4), 263-280.
- 49) Orme, B. K. (2006). Getting started with conjoint analysis. *Madison, WI: Research Publishers LLC*.
- 50) Parducci, A. (1974). Contextual effects: A range-frequency analysis. *Handbook of perception*, 2, 127-141.
- 51) Roest, H., & Rindfleisch, A. (2010). The influence of quality cues and typicality cues on restaurant purchase intention. *Journal of Retailing and Consumer Services*, 17(1), 10-18.
- 52) Ryan, T. P., & Morgan, J. P. (2007). Modern experimental design. *Journal of Statistical Theory and Practice*, 1(3-4), 501-506.
- 53) Saaty, T. L. (1988). What is the analytic hierarchy process? (pp. 109-121). *Springer Berlin Heidelberg*.
- 54) Thomas, L. S. (1988). Multicriteria Decision Making: The Analytic Hierarchy Process.
- 55) Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. *European Journal of Operational Research*, 48(1), 9-26.
- 56) Silk, A. J., & Urban, G. L. (1978). Pre-test-market evaluation of new packaged goods: A model and measurement methodology. *Journal of Marketing Research*, 15(2), 171-191.
- 57) Silva, A., Nayga Jr, R. M., Campbell, B. L., & Park, J. (2007). On the use of valuation mechanisms to measure consumers' willingness to pay for novel products: A comparison of hypothetical and non-hypothetical values. *International Food and Agribusiness Management Review*, 10(2), 165-180.
- 58) Srinivasan, V. (1982). Comments on the role of price in individual utility judgments. *Research in Marketing*, 1, 81-90
- 59) Srinivasan, V. (1988). A Conjunctive-Compensatory Approach to the Self-Explicatin. *Decision Sciences*, 19(2), 295-305..
- 60) Steenkamp, J. B. E., Van Heerde, H. J., & Geyskens, I. (2010). What makes consumers willing to pay a price premium for national brands over private labels? *Journal of Marketing Research*, 47(6), 1011-1024.
- 61) Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science*, 4(3), 199-214.
- 62) Turner, A. D. (1996). A paired-comparison method for interval scaling. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38(2), 362-374.
- 63) Vaidya, O. S., & Kumar, S. (2006). Analytic hierarchy process: An overview of applications. *European Journal of Operational Research*, 169(1), 1-29.
- 64) Vargas, L. G. (1990). An overview of the analytic hierarchy process and its applications. *European Journal of Operational Research*, 48(1), 2-8.
- 65) Veale, R., & Quester, P. (2009). Do consumer expectations match experience? Predicting the influence of price and country of origin on perceptions of product quality. *International Business Review*, 18(2), 134-144.
- 66) Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1), 8-37.
- 67) Wertenbroch, K., & Skiera, B. (2002). Measuring consumers' willingness to pay at the point of purchase. *Journal of Marketing Research*, 39(2), 228-241.

- 68) Wittink, D. R., Vriens, M., & Burhenne, W. (1994). Commercial use of conjoint analysis in Europe: Results and critical reflections. *International journal of Research in Marketing*, 11(1), 41-52.
- 69) Woodworth, R.S., 1978. Experimental psychology. *New York: Holt*.
- 70) Welsh, M. P., & Poe, G. L. (1998). Elicitation effects in contingent valuation: comparisons to a multiple bounded discrete choice approach. *Journal of environmental economics and management*, 36(2), 170-185.