Choice-Based Conjoint Analysis

Eggers, Felix; Sattler, Henrik; Teichert, Thorsten; Völckner, Franziska

Published in:
Handbook of Market Research

DOI:
10.1007/978-3-319-57413-4_23
10.1007/978-3-319-05542-8_23-1

Choice-Based Conjoint Analysis

Felix Eggers, Henrik Sattler, Thorsten Teichert, and
Franziska Völckner

Contents

Abstract

Conjoint analysis is one of the most popular methods to measure preferences of individuals or groups. It determines, for instance, the degree how much consumers like or value specific products, which then leads to a purchase decision. In particular, the method discovers the utilities that (product) attributes add to the

F. Eggers
University of Groningen, Groningen, The Netherlands
e-mail: f.eggers@rug.nl

H. Sattler · T. Teichert
University of Hamburg, Hamburg, Germany
e-mail: henrik.sattler@uni-hamburg.de; thorsten.teichert@uni-hamburg.de

F. Völckner (✉)
Department of Marketing and Brand Management, University of Cologne, Köln, Germany
e-mail: voelckner@wiso.uni-koeln.de

C. Homburg et al. (eds), *Handbook of Market Research*,
https://doi.org/10.1007/978-3-319-57413-4_23

overall utility of a product (or stimuli). Conjoint analysis has emerged from the traditional rating- or ranking-based method in marketing to a general experimental method to study individual's discrete choice behavior with the choice-based conjoint variant. It is therefore not limited to classical applications in marketing, such as new product development, pricing, branding, or market simulations, but can be applied to study research questions from related disciplines, for instance, how marketing managers choose their ad campaign, how managers select internationalization options, why consumers engage in or react to social media, etc. This chapter describes comprehensively the "state-of-the-art" of conjoint analysis and choice-based conjoint experiments and related estimation procedures.

Keywords

Preference measurement · Choice experiments · Conjoint analysis · Conjoint measurement · Tradeoff analysis · Choice-based conjoint · Adaptive conjoint · Utility function · New product development · Revealed preference · Incentive-aligned mechanisms · Willingness-to-pay · Market simulation

Introduction

Assume that an electronics company wants to enter the market for ebook readers. The company has already developed a working prototype with the basic functionality. However, consumers did not yet consider buying this specific product according to a survey, but continue to buy a (more expensive) competitor's product instead. The manufacturer therefore would like to know which attributes of an ebook reader are valued by consumers and which specific attributes they need to improve. Given limited budgets, they can only modify their product in one or two attributes, depending on the manufacturing costs, so that they need to reveal which attributes are most important. Moreover, they would like to know how price-sensitive consumers are and how much they are willing to spend for an ebook reader. Finally, they also need an estimate of the achievable market share to reach the final decision if they should market their product or not.

These questions and related ones can be addressed with preference measurement. The aim of preference measurement is to discover the degree how much consumers like or value (i.e., derive a utility from) specific products, which then leads to a purchase decision. Conjoint analysis, as one of the most popular methods within preference measurement, assumes that products are attribute bundles. Accordingly, an ebook reader is considered as a bundle of screen technology, screen size, screen resolution, storage size, brand name, price, etc. The method tries to discover the utilities that each attribute (and attribute level, respectively) adds to the overall utility of the product by systematically varying specific levels of the attribute. It is a decompositional method, meaning that it elicits consumers' overall utilities for experimentally varied product concepts and then decomposes the overall utility into the attributes' utilities (so-called "partworth utilities" or just "partworths") via statistical procedures. In line with this description, the American Marketing

Association (2015) defines conjoint analysis as a “statistical technique in which respondents’ utilities or valuations of attributes are inferred from the preferences they express for various combinations of these attributes.”

As a result, conjoint analysis provides researchers with a utility function that translates the specific attribute levels of a product into consumers’ preferences. This utility function serves multiple purposes; it can explain consumers’ actual purchase decisions and predict their choices given changes to the product configuration, i.e., modification of attributes. In this regard, it is the basis for a multitude of relevant marketing applications, for example:

– New product development and innovation, e.g., which product concept will be preferred by consumers? (e.g., Page and Rosenbaum 1992; Urban and Hauser 1993)
– Pricing, e.g., how much are consumers willing to pay and how much are improvements in products attributes allowed to cost? (e.g., Miller et al. 2011)
– Branding, e.g., how much value can be attributed to the brand of a product? (e.g., Sattler 2005)
– Market segmentation, e.g., are there different market segments that differ in terms of certain preferred product attributes? (e.g., Teichert 2001b)
– Market scenarios, e.g., what is the effect of a new product entry on the market shares of the incumbents? (e.g., Burmester et al. 2016)

Conjoint analysis is not limited to applications in marketing, but can be generally applied when individuals need to make a decision regarding multiattributive objects. It is also a popular method in other areas, such as transportation (e.g., Hensher 1994), litigation (e.g., Eggers et al. 2016), agriculture (e.g., Lusk and Schroeder 2004), or health economics (e.g., De Bekker-Grob et al. 2012). Due to its broad area of applications, conjoint analysis has advanced to a widely respected method since its introduction into marketing in the 1970s. Overviews of its popularity can be found in Green and Srinivasan (1978, 1990) as well as in empirical studies conducted, for example, by Wittink et al. (1994), Voeth (1999), Sattler (2006), and Orme (2016).

Conjoint methods differ in terms of how the overall utilities are elicited. Traditional approaches use ratings of single product concepts (rating-based conjoint), ratings of pairs of products, or rankings of a selection of products (ranking-based conjoint). Currently, the most popular conjoint approach with over 80% of applications (Orme 2016) is based on choices among several product concepts, i.e., choice-based conjoint (CBC; also termed discrete choice experiments; Haaijer and Wedel 2003; Louviere and Woodworth 1983). Using choices as the dependent variable has become popular because they mimic consumers’ behavior when they are making purchase decisions.

Continuing the example case mentioned above, assume that the manufacturer of the ebook reader is currently producing a black ebook reader with a 6-in. E Ink display and 4 GB storage. They are exploring different options to improve their product, e.g., identified via qualitative research or pretests: (1) increasing the storage from 4 GB to 8 GB, (2) increasing the screen size from 6 to 7 in., or (3) changing the

Table 1 List of potential ebook readers (2^3 design)

Concept	Storage (GB)	Screen size (in.)	Color
1	4	6	Black
2	4	7	Black
3	4	6	White
4	4	7	White
5	8	6	Black
6	8	7	Black
7	8	6	White
8	8	7	White

case color from black to white. Accordingly, there are (2^3) eight different options they could potentially offer, resulting from the different combination of attribute levels (Table 1).

Although one could assume that more storage is better so that 8 GB models are preferred to 4 GB models, this is not necessarily true for screen size since consumers might either value a small (and less bulky) product or a larger (and more readable) screen. There is also no a priori preference order for color. Hence, it is not known beforehand which option would be the most preferred one. Moreover, it might not be profitable to offer an 8 GB model if the increase in preference, and therefore demand, is only marginal and does not justify the additional manufacturing costs. Thus, conjoint analysis is a suitable method to solve this decision problem.

Traditional conjoint analysis (e.g., rating-based conjoint) would present each of the products in Table 1 to a consumer in a survey and ask for his/her preference, e.g., on a rating scale from 0 (“not at all preferred”) to 10 (“very much preferred”). The partworth utilities for the attribute levels can then be derived by using the ratings as a dependent variable in a regression model in which the attribute levels serve as independent variables (e.g., as dummy variables). Although ratings can be considered an acceptable manifestation of preferences, they do not mimic consumers behavior in the marketplace. Moreover, it is often questionable how the ratings can be translated into actual choices (Teichert 2001a).

These issues are among the reasons why CBC approaches have become popular. They offer respondents a selection of product alternatives in a choice set (also called “choice task”) and ask for their most preferred option (Fig. 1). This procedure is repeated across multiple sequential choice sets, each presenting alternatives that are systematically varied by an experimental design. The decisions within a choice set often require a trade-off between attributes. For example, if a consumer prefers larger screens (as in option 1 in Fig. 1) and more storage (as in option 2), she/he needs to determine how important each of these attributes really is in order to reach a decision between option 1 and option 2, while also considering color. These decisions increase the realism of the tasks as trade-off decisions are very often required in the marketplace, e.g., when a higher quality is offered for a higher price. Another element that increases the realism of CBC is that it is possible to include a so-called no-choice option (also termed “none option” or “outside good”), which can be

Which of these ebook readers do you prefer?

Please assume that these two options do not differ in terms of other attributes, i.e., both option have a self-lit E Ink display with 758x1024 pixels resolution, WiFi, and 3 weeks battery life. They both support multiple formats (PDF, EPUB) and connect to major book distributors.

	Option 1	**Option 2**	**Option 3**
Storage:	4 GB	8 GB	I would not buy any of these
Screen size:	7 inch	6 inch	
Color:	White	Black	

Fig. 1 Exemplary choice set of a CBC experiment

chosen if none of the alternatives are acceptable. In this example, the no-choice option could also be termed, e.g., "With these options I would keep reading books on paper," so that a threshold can be identified which indicates the utility that is needed to make consumers switch from traditional books to an ebook reader.

The higher degree of realism of CBC experiments leads to the expectation that CBC exhibits a higher validity compared to traditional, metric conjoint analysis. However, not all studies find significantly better results for CBC compared to traditional conjoint analysis, although the direction of the effects is as expected (Chakraborty et al. 2002; Elrod et al. 1992; Moore 2004; Moore et al. 1998; Vriens et al. 1998). A disadvantage of CBC experiments is that choices among alternatives are nominal and generate less information than, e.g., rating each alternative separately. Therefore, CBC requires collecting a multitude of sequential choice sets, which might invoke respondent fatigue and could serve as an explanation for those findings in which CBC is not predicting significantly better than rating or ranking-based conjoint.

The traditional conjoint approaches (e.g., rating and ranking-based conjoint) and CBC can be classified as static because they do not adapt to the responses that the consumer has given in the survey. To make the information collection more efficient, adaptive procedures dynamically adjust to the preferences of the respondents. They are typically based on a hybrid approach that combines a decompositional and a compositional method. Compositional approaches (e.g., the self explicated method) ask respondents directly about their preference for attribute levels and the relative importance of the attributes, e.g., via rating scales (Srinivasan and Park 1997). This input can then be used as a first estimate of the consumer's preferences in order to show product concepts in the conjoint procedure that are meaningful to the individual respondent or that generate most information about the respondent's preferences. The rating-based Adaptive Conjoint Analysis (ACA, Johnson 1987) and Adaptive CBC (ACBC, Sawtooth 2014) follow this idea. Other adaptive approaches from the machine learning literature dynamically anticipate each respondent's utility based on previous answers, i.e., either ratings (Toubia et al. 2003) or choices (Toubia et al. 2004, 2007). Hybrid individualized two-level CBC (HIT-CBC, Eggers and Sattler 2009) uses a compositional approach in order to ask for the best and worst levels for each attribute and adjusts the CBC part to these two extreme levels only. Thus, it can be seen as a compositional approach in which the attribute importance is derived by a conjoint experiment.

In newer conjoint analysis approaches, respondents interact with each other, following the principles of barter markets (Ding et al. 2009), auctions (Park et al. 2008), or poker games (Toubia et al. 2012). Preferences can then be inferred from these transactions. Figure 2 summarizes the evolution of conjoint analysis approaches.

It should be noted that the above-mentioned example of ebook reader attributes is a very simple case that is used for illustration only. Typically, conjoint studies apply more complex scenarios with more attributes, including price, and additional levels per attribute. Therefore, as an extended example, we will introduce additional attribute levels and a fourth attribute: price. The list of attributes and levels for the extended example is given in Table 2. Because of the popularity of CBC approaches, the remaining chapters will focus on these approaches.

- **Static**
  - Rating-/Ranking-based Conjoint (Srinivasan/Rao 1971)
  - Choice-based Conjoint (CBC) (Louviere/Woodworth 1983)
- **Adaptive**
  - Adaptive Conjoint Analysis (Johnson 1987)
  - Adaptive CBC (Sawtooth 2014)
  - Fast Polyhedral Adaptive Conjoint (Toubia et al. 2003, 2007)
  - Hybrid Individualized Two-Level CBC (Eggers/Sattler 2009)
- **Interactive**
  - Upgrading Auctions (Park/Ding/Rao 2008)
  - Barter Markets (Ding/Park/Bradlow 2009)
  - Conjoint Poker (Toubia et al. 2012)

Fig. 2 Evolution of conjoint analysis approaches

Table 2 Attributes and levels for the extended example

Attribute	Level 1	Level 2	Level 3	Level 4
Storage	4 GB	8 GB	16 GB	n.a.
Screen size	5 in.	6 in.	7 in.	n.a.
Color	Black	White	Silver	n.a.
Price	€79	€99	€119	€139

Model

Conjoint applications assume a (purchase) decision model in which consumer preferences, i.e., utilities, are the central element of the choice process. The assumption is that specific product attributes determine the individual utility evaluations and these, in turn, form the basis for the observed choice behavior (Fig. 3). This requires two interdependent models: a utility model and a choice model, which translates utilities into multinomial choices.

The literature on preference measurement or conjoint-related literature is often equivocal in their terminology. Throughout this chapter, we will use the following terminology (with alternative formulations noted in parentheses): We measure the utility (= preference, need, liking, worth, value) of a consumer (= respondent, individual, subject) for a specific product or service (= alternative, stimulus, object, option, profile) that consists of different attributes (= factors, dimensions), each having specific attribute levels (= characteristics, features).

Utility Model

The basis for the utility model in a choice context is random utility theory (RUT), which states that the overall utility U of consumer c for a product i is a latent construct that includes a systematic component V and an error component e, i.e., $U_{ci} = V_{ci} + e_{ci}$ (McFadden 1981; Walker and Ben-Akiva 2002). The stochastic error term catches all effects that are not accounted for and can include, e.g., respondent fatigue, omitted variables, biases in the data collection, or unaccounted heterogeneity (Louviere and Woodworth 1983).

The theory assumes that a consumer chooses the product from a set of alternatives that exhibits the highest utility. Since the overall utility is influenced by a stochastic component, it is only possible to state a probability that this consumer would choose the product. Consequently, the probability p that a consumer chooses product i from a set of products $S = \{i, j\}$ is (Train 2009):

$$p_i = p\left(U_i > U_j\right) = p\left(V_i - V_j > e_j - e_i\right) \tag{1}$$

According to Eq. (1) a consumer is more likely to choose product i if the utility of i is larger than the utility of j. This requires that there is a positive residual from the difference in systematic utilities and that this residual exceeds the influence of error. Consequently, only differences in product attributes are considered, e.g., if consumers need to choose between two ebook readers and both devices are black then

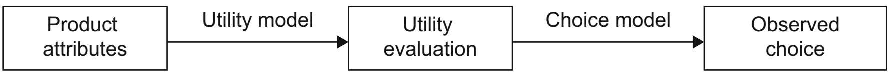


Fig. 3 Elements of a purchase decision model

color does not affect the decision. Generally, any constant value can be added to the utility functions and it will not affect the outcome, which is why choice-based utilities are interval-scaled and choice models do not have a general regression constant (constants, if any, need to be alternative-specific).

The systematic utility V represents the function that translates the product attributes and their levels into partworth utilities. The estimated utility V_i for a product i with N attributes can be divided into two subfunctions ψ and f_n as follows (Teichert 2001a):

$$V_i = \Psi[f_1(v_{1i}), f_2(v_{2i}), \ldots, f_N(v_{Ni})] \tag{2}$$

with

v_{ni}: Partworth utility of attribute n in product i, $n = 1, 2, \ldots, N$
f_n: Evaluation function of attribute n, $n = 1, 2, \ldots, N$
ψ: Function to combine partworth utilities across attributes

Evaluation Function for Attribute Levels

The function f_n in Eq. 2 describes how levels of attribute n are evaluated. The basic idea is that at least one attribute level represents the ideal point for the consumer (or at least the most preferred level from the available attribute levels). Differences to this ideal point lead to a loss in utility. Figure 4 depicts three potential functional forms.

The vector model assumes that increasing (decreasing) the attribute level leads to a proportional positive (negative) effect in utility. Hence, the ideal point is positive (negative) infinity. This model would be appropriate when assuming, e.g., that increasing the screen size of an ebook reader from 5 to 6 in. leads to the same positive utility difference as upgrading the screen from 6 to 7 in. The vector model uses the actual numeric values of the attributes and just one utility parameter to represent the partworth utility:

$$v_{in} = \beta_n * X_{inm} \tag{3}$$

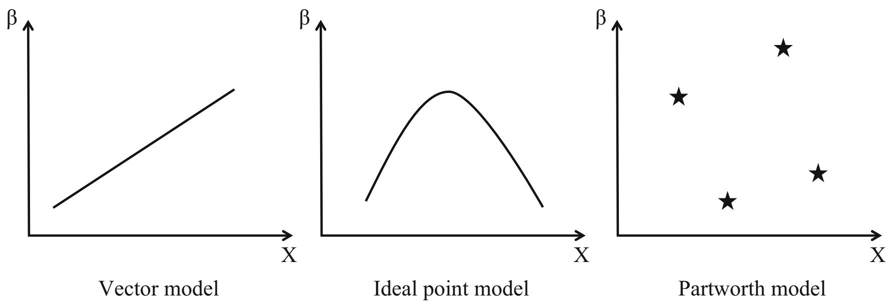


Fig. 4 Alternative functional forms for the evaluation of attribute levels

with,

v_{in}: partworth utility for attribute n in product i
β_n: utility vector for attribute n
X_{inm}: numeric value of level m of attribute n in product i

The ideal point model does not assume a linear slope of the utility function as the vector model but assumes diminishing (or increasing) marginal utilities. For example, although consumers might in general prefer larger screens for an ebook reader, very large sizes will become impractical so that utilities will decrease again when increasing the size from an (individually perceived) ideal point further. Likewise, when an ebook reader already has a very large storage, it can be expected that increasing the storage further leads to a diminishing marginal utility for the consumer. The ideal point model thus considers not only the numeric value of the attribute level, e.g., its screen size, but also its squared term:

$$v_{in} = \beta_{n1} * X_{inm} + \beta_{n2} * X^2{}_{inm} \tag{4}$$

with,

v_{in}: partworth utility for attribute n in product i
β_{n1}: utility vector for attribute n
β_{n2}: utility vector for the squared value of attribute n
X_{inm}: numeric value of level m of attribute n in product i

The partworth model estimates separate partworth utilities for each level of the attribute, i.e., there is no assumed functional relationship between the attribute levels. This model is required for qualitative, nominal attributes, e.g., color, but can also be applied to quantitative, numeric attributes. If the choice sets include a no-choice option, this option is also represented by a separate partworth that measures the attractiveness of not choosing any of the alternatives. The partworth model is typically based on dummy-coding (or effect-coding) techniques, which requires $M-1$ variables to represent an attribute with M levels:

$$v_{in} = \sum_{m=1}^{M-1} \beta_{nm} * X_{inm} \tag{5}$$

with,

v_{in}: partworth utility for attribute n in product i
β_{nm}: partworth utility for level m of attribute n
X_{inm}: dummy variable with value 1 if product i features level m of attribute n, otherwise 0

Regarding the number of parameters that these models require for the estimation, the vector model is the most parsimonious as it only uses one parameter per attribute.

The ideal point model is based on two parameters. The partworth model requires setting one attribute level as the reference level, which is left out of the estimation so that it requires $M - 1$ parameters.

The partworth model can be considered conservative since it does not require a prior specification or theory about the slope of the partworth utility function. If more than two attribute levels are present, it uses the most number of parameters and therefore provides the best model fit (by sacrificing degrees of freedom). It is therefore not surprising that the partworth model is predominantly used in conjoint analysis and is partly also considered as a constitutive element (Shocker and Srinivasan 1973).

Function to Combine Partworth Utilities Across Attributes

The function ψ in Eq. 2 determines how to combine partworth utilities across attributes. Conjoint analysis assumes a compensatory utility model. In a linear additive utility model, the overall systematic utility V_i of a product i is the sum of the partworth utilities v_{in} of its attributes $n = 1, \ldots, N$:

$$V_i = \sum_{n=1}^{N} v_{in} \tag{6}$$

Complex functions can be modeled as extension to this base model, e.g., interaction effects between attributes. Interaction effects occur when the utility evaluation of one attribute level depends on the level of another attribute. For example, consumers might prefer a white color for ebook readers with large screens but black for readers with smaller screens.

Interaction effects can be modeled as additional effects in the linear additive base model by including separate partworth utilities for the cross product of two attributes. The overall utility for a product is then represented as the sum of the partworth utilities of both the main effects and the interaction effects:

$$V_i = \sum_{n=1}^{N} v_{in} + \sum_{m=1}^{M-1} \sum_{m'=1}^{M'-1} \beta^{IA}_{nm,n'm'} * X_{inm} * X_{in'm'} \tag{7}$$

with,

$\beta^{IA}_{nm;n'm'}$: Interaction effect between level m of attribute n and level m' of attribute n'; $m = 1, 2, \ldots, M$; $m' = 1, 2, \ldots, M'$

$X_{inm}; X_{in'm'}$: Dummy variable with value 1 if product i features level m (m') of attribute n (n'), otherwise 0

Interaction effects increase the complexity of a model. For this reason, they are predominantly added if theory or prior assumptions about them exist. However, being able to measure interaction effects with conjoint analysis is a major advantage compared to other survey techniques, e.g., compositional approaches.

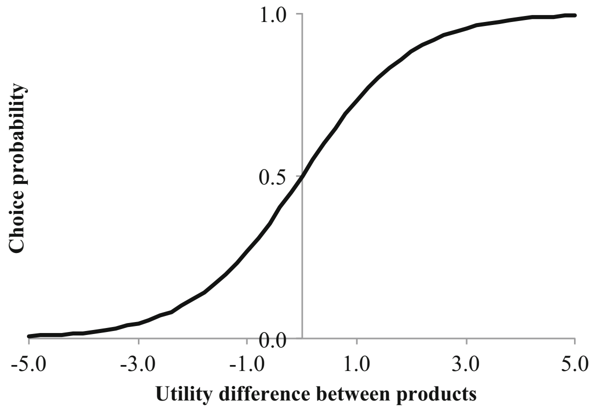


Fig. 5 S-shaped function of the multinomial logit model

Choice Model

Choice models can be differentiated according to the assumptions about the stochastic error component (see Train 2009 for an overview). In most applications, the error is assumed to be independent and identically distributed (iid) as extreme value type, i.e., Gumbel. This assumption leads to a logistic distribution of the differences of error terms and the multinomial logit (MNL) model (McFadden 1981; Hensher and Johnson 1981; Louviere et al. 2000). Accordingly, choosing an object i from a choice set with S alternatives is represented by the MNL model in terms of choice probabilities p:

$$p(i|S) = \frac{\exp(V_i)}{\sum_{j \in S} \exp(V_j)} \tag{8}$$

The MNL model results in an S-shaped relationship between utility difference and choice probability (Fig. 5).

An alternative to the Gumbel distribution is the assumption of a normal distribution of the error term, which results in a multinomial probit model (Haaijer et al. 1998). The probit model requires multiple integrals and complex estimation procedures. Because of the compact form of the logit function (see Eq. 8), the MNL model is predominantly applied in CBC analyses (Haaijer and Wedel 2003).

Procedure for Conducting Discrete Choice Experiments

Identification of Attributes and Attribute Levels

The prerequisite – and most relevant step – for conducting conjoint analyses is to identify the relevant determinants of consumers' choices, i.e., product attributes and their levels. The selection of attributes and levels should reflect the products on the marketplace and should affect consumers' preferences. Otherwise, the validity of the model can be questioned. In general, the selection of attributes has to fulfill the following requirements (Green and Srinivasan 1978, 1990; Orme 2002):

– Attributes should be relevant, i.e., they should influence consumers' utility. In order to identify relevant attributes qualitative surveys, e.g., focus groups or depth interviews can be used.
– Attributes should discriminate, i.e., they should be able to differentiate between the competitive offerings on the marketplace.
– The number of attributes should be manageable. CBC experiments typically use less than seven attributes. Using more attributes greatly increases the complexity of the experimental design and requires high cognitive capabilities of the respondents.
– Attributes should not be interrelated, i.e., they should measure independent aspects of the product. If attributes are interrelated, then certain combinations might be highly unrealistic and confusing to the respondents. However, if, e.g., higher storages typically go along with higher prices, it is possible to consider these attributes as independent and analyze "what-if" scenarios. It should be noted that this requirement does not preclude potential interaction effects, i.e., although the attributes are independent, it does not mean that the preferences for them are as well.

After setting the attributes, their levels need to be determined. Regarding the type and number of levels, the following requirements should be considered (Green and Srinivasan 1978, 1990; Orme 2002; Teichert 2001a):

– The levels should span a range that is larger than in reality, but not substantially, in order to be able to cover potential future scenarios.
– Levels that have an ambiguous meaning should be avoided. For example, instead of using levels "large" and "small" for screen size, it is better to use specific values because they are free from interpretation. Moreover, specific values allow using a vector or ideal point model for estimation.
– The number of levels should be kept low because the complexity of the experimental design will increase exponentially with more levels. Consider the example in Table 1 with $2^3 = 8$ combinations. If three levels per attribute were used instead there are already $3^3 = 27$ potential options. Conjoint experiments can consider complex designs, however, most applications use an average of three to four levels per attribute.
– When setting the number of attribute levels, it should also be considered if the linearity or nonlinearity of the utility function (e.g., an ideal point model) should be tested, which then requires at least three levels. For testing interaction effects, it would be preferable (but not required) to use just two levels in order to keep the number of interaction effect parameters low.
– The number of levels should be balanced across attributes. Otherwise, the number-of-levels effect can occur, which leads to an artificially higher relevance of attributes that have more levels (Eggers and Sattler 2009; Verlegh et al. 2002).

– Levels should be generally acceptable. Unacceptable levels would otherwise invalidate the assumed compensatory utility model.
– Attribute levels are assumed to be mutually exclusive. For example, if an attribute "extra features" is added to the ebook reader setup with the levels "waterproof" and "integrated music player," the reader can only have one of these levels. If it is also interesting for the researcher to analyze preferences for both features in combination, this combination should be added as a separate level (an alternative would be to define each extra as a separate attribute with the levels "yes" and "no").

Creating the Experimental Design

The experimental design determines which combinations of attribute levels are presented to the respondent as stimuli (factorial design) and how these stimuli are allocated to choice sets (choice design). It represents the independent variable matrix for the analysis. To estimate the main effects of the attributes – and potentially interaction effects between them – the experimental design needs to make sure that these effects can be identified.

Criteria to evaluate the efficiency of an experimental design are (Huber and Zwerina 1996):

– Balance, i.e., each attribute level is presented an equal number of times
– Orthogonality, i.e., attribute levels are uncorrelated
– Minimal overlap, i.e., alternatives within a choice set are maximally different
– Utility balance, i.e., alternatives within a choice set should be equally attractive so that there should not be dominated or dominating alternatives

Balance and orthogonality refer to the factorial design, while minimal overlap and utility balance relate to the choice design.

Factorial Design

The set of all potential stimuli, i.e., every combination of attribute levels, leads to a full factorial. With N attributes and M_1 levels for attribute 1, M_2 levels of attribute 2, and M_N levels of attribute N, the size of the full factorial consists of all permutations $M_1 * M_2 * \ldots * M_N$. Table 1 shows a full factorial of the 2^3 design. Full factorials are always balanced, i.e., the attribute levels occur an equal number of times (here, four times), and orthogonal, i.e., each pair of attribute levels is balanced (here, each pair occurs twice).

A full factorial is only required if all main effects and all potential interaction effects should be estimated. The 2^3 design with three binary attributes A, B, C allows to estimate the three main effects, the three two-factor interaction effects (A*B, A*C, B*C), as well as the three-factor interaction (A*B*C). This is demonstrated in Table 2, in which the attribute levels are effect-coded (first level = 1, second level = −1). The interaction levels result from multiplying the levels of the underlying main effect attributes. As can be seen, the resulting interaction levels are not identical to any other column, i.e., are independent, and are also balanced and orthogonal so that they can be identified.

Since the full factorial increases exponentially when more attributes and/or more attribute levels are added, its size quickly becomes hard to handle in an experimental survey. For example, the extended example with three three-level attributes and one four-level attribute consists of 3^3 * 4 = 108 potential alternatives. Moreover, very often three-factor interaction effects can be neglected and not all two-factor interaction effects may be required. In general, smaller factorials, i.e., fractional factorials, still allow estimating main effects and selected interaction effects (Addelman 1962).

The idea of creating a fractional factorial design is demonstrated with an example. Consider that a fourth binary attribute D would be added to the simple example in Table 3. The full factorial would then increase to 2^4 = 16 stimuli. A fractional design assumes that at least one of the interaction effects between the attributes A, B, and C would be zero so that it can be replaced with the main effect of D, e.g., D = A*B, i.e., each level of the interaction between A and B becomes the new level of D. The fractional factorial then consists of the 8 entries in Table 2 and columns A, B, C, as well as D = AB. The factorial was reduced to 8 stimuli, i.e., by 50% compared to the full factorial. Nevertheless, it is still able to identify all main effects, i.e., the design is still balanced and orthogonal. As a downside, however, the interaction effect between A and B cannot be estimated as it is confounded with the main effect of D.

Fractional factorials are documented for the most common experimental designs (e.g., Sloan 2015) or can be generated via software (e.g., SAS or SPSS). The efficiency of the fractional design can be tested easily by checking the correlation matrix of all assumed main and interaction effects. If there are no or only minor correlations, then the design is orthogonal and the parameters can be identified without bias.

For traditional rating- or ranking-based conjoint procedures, it is sufficient to evaluate the factorial design. CBC methods require an additional step of allocating alternatives of the factorial design to specific choice sets, i.e., to evaluate the choice design.

Table 3 Main and interaction effects of a full factorial 2^3 design

	Main effect			Interaction effects			
Stimulus	A	B	C	AB	AC	BC	ABC
1	−1	−1	−1	1	1	1	−1
2	−1	−1	1	1	−1	−1	1
3	−1	1	−1	−1	1	−1	1
4	−1	1	1	−1	−1	1	−1
5	1	−1	−1	−1	−1	1	1
6	1	−1	1	−1	1	−1	−1
7	1	1	−1	1	−1	−1	−1
8	1	1	1	1	1	1	1

Choice Design

Choice experiments require that the factorial is subdivided into choice sets with a selection of alternatives. Creating an optimal choice design involves complex algorithms based on combinatorics. For example, even with the simple example and a 2^3 full factorial, there are $\binom{8}{2} = 28$ different choice sets with two alternatives. The complexity increases with the size of the factorial, e.g., in the extended example there would be $\binom{108}{3} = 205{,}156$ potential choice sets of size three. The challenge lies in selecting those choice sets that provide the most information about the respondents' preferences. The efficiency criteria minimal overlap and utility balance help reducing the size of the list of potential choice sets (Huber and Zwerina 1996).

Minimal overlap requires that the alternatives within a choice set are maximally different, i.e., have different attribute levels (Sawtooth 1999). It is based on the idea that an attribute that exhibits the same level for each alternative within a set does not affect the choice (see Eq. 1). A choice design with minimal overlap can be created for the simple example when the first four entries in the full factorial in Table 1 are coupled with their fold-over, i.e., opposite level. Accordingly, concept 1 (4 GB, 6 in., black) would be coupled with concept 8 (8 GB, 7 in., white) to create one choice set; concept 2 would be coupled with concept 7, etc., so that in total four choice sets with minimal overlap are created.

The idea of selecting choice sets that are utility balanced is that alternatives are allocated to a choice set that are equally attractive (Huber and Zwerina 1996). Contrarily, a choice set that features a dominating or dominated alternative provides no new knowledge since the choice can be anticipated. However, dominating alternatives can only be identified if there is a priori knowledge about the respondents' preference structure or if respondents' preferences are anticipated during the experiment with adaptive conjoint approaches (see above).

Because of the complexity of creating an optimal choice design, computer algorithms are recommended. For example, SAS or Sawtooth offer algorithms to create optimal choice designs and analyze their efficiency.

Decision Parameters

Relevant decision parameters for the experimental design also concern the number of stimuli per choice set and the number of choice sets.

Each choice task should be manageable for the respondent, which favors showing only a few alternatives per set (Batsell and Louviere 1991). On the other hand, more alternatives increase the information of each choice. Therefore, two to five stimuli per choice set are most common (Meissner et al. 2016). Using eye-tracking data, Meissner et al. (2016) show that the number of alternatives also affects search patterns. It is therefore advisable to use a choice set size that is similar to the typical size of a consideration set when consumers make purchase decisions. In product categories in which consumers frequently have to choose from a multitude of alternatives, e.g., toothpaste in supermarkets, choice sets could also include a larger number of alternatives (Hartmann 2004). The selection of the number of alternatives should also consider the number of attribute levels since using a number of alternatives that is a subset of the number of levels provides statistical benefits (Zeithammer and Lenk 2009).

Apart from the number of alternatives per choice set, the number of choice sets needs to be considered when selecting an optimal design. More choice sets lead to a higher reliability of the parameters. However, from a consumer perspective, more choice sets induce fatigue so that respondents tend to make more errors or even switch their decision strategy, e.g., focusing more on the price attribute (Johnson and Orme 1996), which is counterproductive. Consistently, results concerning the predictive validity depending on the number of choice sets indicate that the marginal benefit of additional choice sets declines (Sattler et al. 2004; Teichert 2001a). A review of articles published in the *Journal of Marketing Research* between 2000 and 2017 shows that most researchers make a compromise between statistical reliability and consumer fatigue so that most applications (14 out of 42) have used 11–15 sets. Slightly fewer studies (13 out of 42) have used ten sets or less. The number of applications decreases with more choice sets, i.e., nine studies used 16–20 choice sets, five applications 21–25 sets, and one study more than 25.

Implementation into Questionnaire

The implementation of the CBC experiment into a questionnaire requires decisions regarding the presentation of stimuli, integration of a no-choice option, collecting additional choices per choice set, applying incentive alignment mechanisms, and adding holdout choice sets.

Presentation of Stimuli

Most CBC interviews are computer-based since they facilitate handling complex experimental designs. Moreover, having more than two alternatives per choice set puts high cognitive burden on respondents, e.g., when described via telephone interviews. Computer-based interviews are beneficial because they allow implementing attribute levels or overall stimuli as multimedia information. Instead of using text only it is possible to depict the size of the ebook reader screens as a pictogram or to show actual ebook readers in different colors. When certain functionalities, e.g., page-turn effects, are included as attributes, these could be showcased with instructional videos (e.g., following the idea of information acceleration, Urban et al. 1996). Eggers et al. (2016) demonstrate that the more realistic the experiment can be made compared to what consumers see in the marketplace, i.e., investing in “craft,” the higher is the validity of the results, which might also change the managerial implications from the results compared to studies that rely on defaults, e.g., text-only descriptions of the stimuli.

No-Choice Option

An advantage of CBC experiments compared to metric (rating or ranking-based) conjoint analyses is that respondents can indicate that they prefer none of the presented alternative. This none (or no-choice) option increases the realism since it does not force a decision if the alternatives are unacceptable so that consumers would not buy any of them or switch stores in reality (Haaijer et al. 2001). Recent approaches suggest asking for the no-choice option separately, i.e., sequentially after each choice set ("dual response none"; Brazell et al. 2006). In the dual response procedure respondents are first asked to select the most preferred option (excluding no-choice) in a forced-choice task and, sequentially, whether they would purchase the selected product concept in a second step (Brazell et al. 2006; Wlömert and Eggers 2016).

This procedure allows observing the preferred alternative even if it is not acceptable to be purchased. At the same time, consumers have no possibility to opt out of difficult decisions. Moreover, Wlömert and Eggers (2016) show that the increased salience of the no-choice option leads to more realistic predictions of adoption shares.

The no-choice option plays a central role when calculating (absolute) willingness-to-pay (see section "Market Simulations"). Implications from these analyses are limited if consumers show extreme response behavior and never or always choose the none option. To avoid these extremes, Gensler et al. (2012) present an adaptive approach that dynamically adjusts the price levels downwards whenever the respondent selected the no-choice option and upwards whenever the respondent selected an alternative. Schlereth and Skiera (2016) address this issue by proposing a separated adaptive dual response (SADR) procedure. They adjust the dual response procedure so that the forced choice and purchase question are not presented within the same task but are separated into sequential blocks. Presenting the block of forced choices first allows them to approximate the utility of the alternatives and adaptively select fewer, but more informative alternatives (not necessariliy the alternatives selected in the forced choices) in the purchase questions thereafter.

Collecting Additional Choices per Choice Set

Recently, it was suggested to ask not only for the best option but also for the worst option in a so-called best-worst scaling (or MaxDiff) approach (Louviere et al. 2015; Sawtooth 2013). By assuming that worst choices are reversed best choices, both decisions measure the same construct, i.e., preferences. Stated differently, if β_{nb} represents the partworth utility for attribute n based on best choices and β_{nw} is the partworth utility for the same attribute based on worst choices then it can be assumed that $\beta_{nb} = -\beta_{nw}$. The choices can then be used to make the estimation more reliable since twice as many observations exist. Collecting more choices per set is not limited to best and worst decisions only. More choices can be used as separate dependent variables in order to explore different aspects of consumers' preferences. An additional choice can be, e.g., "Which of these ebook readers would you buy for your partner?" which might explore consumers' gift giving behavior. In a study by Kraus et al. (2015), the authors collected additional choices per set to analyze managers' perception of risk and success of different internationalization strategies.

Figure 6 shows an example of a choice set that includes best and worst choices and a dual response no-choice option.

Which of these ebook readers is your most preferred option and which option is the least attractive?

Please assume that these options do not differ in terms of other attributes, i.e., all options have a self-lit E Ink display with 758x1024 pixels resolution, WiFi, and 3 weeks battery life. They support multiple formats (PDF, EPUB) and connect to major book distributors.

	Option 1	Option 2	Option 3
Storage:	4 GB	8 GB	16 GB
Screen size:	6 inch	7 inch	5 inch
Color:	Silver	Black	White
Prize:	€99	€119	€139
Best option:	◯	◯	◯
Worst option:	◯	◯	◯

Would you actually buy your most preferred option if it was available?

◯ Yes

◯ No

Fig. 6 Choice set with best and worst choices and dual response no-choice option

Incentive Alignment

Ding et al. (2005) introduced incentive alignment mechanisms to conjoint analysis. The basic idea of incentive-aligned (IA) mechanisms is to attenuate hypothetical bias by influencing the type of reward that is provided to respondents. Specifically, the reward is linked to the preferences the respondent expresses during the data collection.

Ding et al. (2005) implemented the IA mechanism by rewarding the respondent with the alternative that she/he selected in a randomly selected choice task (including the no-choice option). In this way, each choice might constitute the potential reward so that respondents are motivated to answer truthfully. If the study features a price attribute, then respondents are required to actually purchase the product for the price shown. Payment is typically achieved by providing the respondents with a budget. If the respondent selected the no-choice option, she/he gets the full budget as a monetary reward. If she/he selected a product for a price €X, she/he will get the actual product plus the remaining change (i.e., initial budget minus €X).

Ding (2007) proposed an alternative IA approach in which respondents are informed before completing the choice tasks that their choices will be used to infer their willingness-to-pay (WTP) for one specific product concept (see sections "Willingness-to-Pay" and "Market Simulations" for details about calculating WTP). Under this WTP-based mechanism, incentive alignment is achieved by obliging participants to purchase this specific product concept at a randomly drawn price if this random price is less or equal to the WTP inferred from the CBC experiment. This approach integrates the incentive compatible Becker-Degroot-Marschak (BDM) auction procedure (Becker et al. 1964, see also Wertenbroch and Skiera 2002) with CBC analysis. Ding (2007) shows theoretically that truthful answers constitute the Bayesian Nash equilibrium for participants in such applications as long as the respondents do not know the configuration of the product that is used as a reward prior to the study.

Dong et al. (2010) introduced and validated a third variant of IA conjoint experiments which involves predicting a rank ordering of the possible rewards based on estimated preferences. Eventually, the reward that is predicted to be ranked first is given to the respondent. Again, respondents are motivated to answer truthfully and keep the impact of error small in order to be rewarded with their most preferred product.

It has been shown that incentive-aligned (IA) data collection procedures substantially increase the predictive performance of conjoint choice experiments compared with traditional CBC analysis (Ding 2007; Ding et al. 2005; Dong et al. 2010) so that their application is recommended. However, one drawback of incentive alignment is that their application is limited to contexts where at least one concept of the research object can be rewarded after the experiment. This may not be feasible in many instances, for example, when the research object is an innovative product and not yet available on the market.

Holdout Choice Sets

A holdout choice set is a choice task that mimics a regular choice set but that is not used in the estimation. The answers given in the holdout choice set provide a benchmark for the (internal) predictive validity of the estimation results. The better the preference estimates are able to predict the actual choices made in the holdout sets the higher the predictive validity. Validity can be assessed with different measures. The hit rate compares on an individual level if the predicted most preferred alternative based on the estimates equals the alternative actually chosen in the holdout set, i.e., a hit meaning a correct prediction. The hit rate is then the mean value across all respondents. The mean absolute error (MAE), as an alternative measure among others, considers the absolute differences between predicted and actual choice shares for each alternative in the holdout set (e.g., Moore et al. 1998).

Estimation

Since choices from choice sets typically do not provide enough information to estimate reliable utilities at the individual level, they require some level of aggregation (see Frischknecht et al. 2014 for an alternative approach). The estimation procedure described here is based on the maximum likelihood procedure. It aggregates all choices from all respondents and produces one set of utilities that represent all consumers, i.e., it neglects consumer heterogeneity (see section "Advanced Estimation Techniques" for advanced estimation procedures without this assumption).

We will use the MNL model for describing the estimation in more detail. The estimates are based on the extended ebook reader example. The (simulated) data are based on 200 respondents who answered 10 choice sets, each showing three product alternatives plus a no-choice option.

Coding

The estimation of partworth utilities requires transforming the attribute levels according to a dummy (or effect) coding technique. When applying a partworth utility model to an attribute with M levels, M − 1 dummy-coded variables are needed to represent this attribute in the estimation. Each variable represents one attribute level and can take the values 1 or 0 depending on whether the attribute level was shown or not. The M^{th} attribute level (or any other level) is left out since it can be expressed as a linear combination of the other variables and cannot be estimated separately. The partworth utility of this reference level is set to 0. The partworth utilities of the remaining attribute levels need to be interpreted in relation to this level. Thus, it matters for the interpretation which level represents the reference.

Conjoint experiments are frequently coded using effect-coding. Effect-coded variables (Louviere et al. 2000), as an alternative to dummy-coding, are zero-centered so that the sum of partworth utilities across all levels of the attribute is zero, i.e., positive partworth utilities indicate higher preferences for that level compared to the average partworth utility across all levels of the attribute. Therefore, positive or negative values do not necessarily mean that these levels are perceived as positive or negative on an absolute level but only compared to the mean of the levels that were included in the experimental design. The reference level, which is left out of the estimation, can be recovered by calculating the partworth utility that is needed so that the average across all utilities is zero. Effect-coding therefore provides a partworth utility value for each attribute level, and it is irrelevant which level is set as the reference.

Effect-coding can be accomplished by setting the reference level to −1, instead of 0 as in dummy coding. Table 4 shows an example of effect-coding two attributes with M = 3 and M = 4 levels. Figure 7 shows an excerpt of the first two choice sets from the ebook reader dataset. In this dataset, each alternative (indicated by Alt_id) is represented by one row such that four rows represent one choice set (indicated by Set_id). The none option is included as one of the alternatives, which is represented by the None variable. The columns in dark grey show the numeric values for screen size, storage, and price, and text information for color. Effect-coding (columns in light grey) needs two parameters each for the attributes storage, screen size, and color, and three parameters for the effect-coded prices. This means that a partworth model requires ten parameters in total, i.e., nine parameters for the effect-coded variables and one variable for the none option (here, dummy coded). The column Selected is a dummy coded variable that shows which alternative was chosen in each choice set. It serves as the dependent variable in the estimation model.

Table 4 Effect-coding of attribute levels

	Effect-coded variables for M = 3		Effect-coded variables for M = 4		
Level	X_1	X_2	X_1	X_2	X_3
1	1	0	1	0	0
2	0	1	0	1	0
3	−1	−1	0	0	1
4			−1	−1	−1

Resp_id	Set_id	Alt_id	Selected	None	Storage	Screen.size	Color	Price	Storage_4GB	Storage_8GB	Screen.size_5inch	Screen.size_6inch	Color_black	Color_white	Price_79	Price_99	Price_119
1	1	1	0	0	4	7	Silver	119	1	0	–1	–1	–1	–1	0	0	1
1	1	2	1	0	16	5	White	79	–1	–1	1	0	0	1	1	0	0
1	1	3	0	0	8	6	Black	99	0	1	0	1	1	0	0	1	0
1	1	4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	1	0	0	8	5	Silver	139	0	1	1	0	–1	–1	–1	–1	–1
1	2	2	0	0	16	6	White	79	–1	–1	0	1	0	1	1	0	0
1	2	3	0	0	4	7	Black 0	119	1	0	–1	–1	1	0	0	0	1
1	2	4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 7 Excerpt from the ebook reader dataset

Maximum Likelihood Estimation

Applying OLS procedures for the estimation is not appropriate because CBC analyses provide nominal data. The estimation of the MNL model therefore relies on maximum likelihood procedures. In aggregate-level analyses, all respondents are pooled to estimate one set of partworth utilities for the entire sample (Louviere and Woodworth 1983; Sawtooth 1999).

The maximum likelihood procedure aims at finding the set of partworth utilities that best represents the observed choices. The likelihood function L results from multiplying the MNL probabilities as shown in Eq. (8) across all choice sets $t = 1, 2, \ldots, T$ and – in the aggregate-level estimation – across all respondents $c = 1, 2, \ldots, C$ (Louviere et al. 2000):

$$L = \prod_{c}^{C} \prod_{t}^{T} p(i_{tc} | S_{tc}) \tag{9}$$

with,

i_{tc} = chosen alternative in choice set t by respondent c
S_{tc} = alternatives in choice set t presented to respondent c

The parameters can be found by maximizing the function subject to the partworth utilities, i.e., $\frac{\partial L}{\partial \beta} = 0$.

The likelihood function lies in the interval [0, 1] and expresses the aggreggate probability to observe the choice data given the set of estimated partworth utilities. However, the minimum of zero is only a theoretical value as choosing randomly between the choice options, i.e., assuming that all betas are zero, would yield a probability of 1/S, with S being the number of alternatives in the choice set. For example, choosing randomly between three ebook readers and the no-choice option would give a probability of 1/4 that the choice matches the respondents preferred option. The lowest logical value of the likelihood function is therefore (1/S)^(T*C). Since this value is very close to zero, the optimization of the function is typically based on the logarithm, i.e., log-likelihood function (Louviere et al. 2000). The lowest value, and the benchmark to assess the model fit, then is T*C*log(1/S), e.g., for the ebook reader case with 10 choice sets with four alternatives and 200 respondents: 10*200*log(1/4) = −2772.6. The estimation model should exceed this value significantly, i.e., have a log-likelihood value that is less negative (closer to zero), because otherwise the partworth utilities would not predict choices better than a random, NULL model.

Estimating the partworth utilities based on the ebook reader example yields a log-likelihood value of −2277.8. To test if the difference in log-likelihood between the NULL model and the estimated model is significant, a likelihood ratio test can be applied. The test statistic is $\chi^2 = 2 * (LL_1 - LL_0)$, with LL_1 representing the log-likelihood of the estimated model and LL_0 the log-likelihood value of the NULL model. This test statistic is distributed χ-squared with degrees of freedom (df) equal to the difference in the number of parameters between both models. In this case, χ^2 is

$2 * (-2277.8 - (-2772.6)) = 989.6$, with df = 10. This test is highly significant ($p < 0.001$), i.e., the estimated model predicts significantly better than the NULL model.

Another measure to assess the goodness of fit is the Pseudo-R^2 or McFadden's $R^2 = 1 - (LL_1/LL_0)$. For the ebook reader example, it is: $R^2 = 1 - (-2277.8/-2772.6) = 0.178$. McFadden's R^2 can be adjusted according to the number of parameters, i.e., $1 - ((LL_1 - npar)/LL_0)$, with npar being the number of parameters. This R^2 value has a different interpretation than in linear regression models. Typically, values exceeding 0.2–0.4 are considered acceptable. Although the ebook reader model is significantly different from the NULL model, its fit relative to this benchmark is not exceeding the threshold of 0.2. A potential explanation for this low fit is that consumers likely have heterogeneous preferences, e.g., towards screen size or color, which are not acknowledged in the aggregate model and therefore increase the error term.

The estimated partworth utilities are depicted in Table 5 (see "Appendix" for the corresponding R code). The partworth utilities for the attribute levels are effect-coded, which can be seen by checking that the sum across the betas is zero. The betas for storage and price show face validity as increasing the storage (price) yields higher (lower) utilities. There is no such trend regarding screen size as 6-in. models have the highest utility, followed by 5-in. models and 7-in. screens. White ebook readers are more preferred than black and silver models.

The no-choice option was dummy coded in this case, with "no-choice" equal to one and "not the no-choice" equal to zero. As can be seen, not choosing one of the

Table 5 Estimated partworth utilities based on the aggregate-level model

Attributes	Beta	Standard error	t-value	Attribute importance
Storage				21.6%
4 GB	−0.389	0.042	−9.323	
8 GB	−0.051	0.039	−1.322	
16 GB	0.440	0.036	12.143	
Screen size				22.0%
5 in.	−0.049	0.039	−1.274	
6 in.	0.446	0.036	12.352	
7 in.	−0.397	0.042	−9.528	
Color				12.5%
Black	−0.002	0.038	−0.059	
White	0.240	0.037	6.547	
Silver	−0.238	0.040	−5.952	
Price				43.9%
€79	0.840	0.045	18.502	
€99	0.286	0.047	6.103	
€119	−0.284	0.053	−5.416	
€139	−0.842	0.063	−13.447	
No-choice				
	−0.532	0.069	−7.749	

ebook readers shows a negative partworth utility so that on average (i.e., with all attributes at their mean utility of zero), choosing one of the ebook readers provides a higher utility and is therefore more likely than choosing none.

The partworth utilities can be transformed to be more accessible for managerial use compared to the rather abstract units of utility. Three transformations shall be elaborated subsequently: relative attribute importances, willingness-to-pay measures, and calculation of purchase probabilities within market simulations.

Relative Attribute Importance

The attribute importance w_n of an attribute n can be calculated based on the relative range of the partworth utilities, i.e., the difference between the most and least preferred attribute levels related to the sum of ranges across all attributes:

$$w_n = \frac{\max(\beta_n) - \min(\beta_n)}{\sum_{i=1}^{N} (\max(\beta_i) - \min(\beta_i))} \tag{10}$$

For example, storage exhibits a range of 0.829 (=0.440 − (−0.389)). The sum of all attribute ranges is 3.832. The relative importance of storage is therefore 0.829/3.832 = 21.6%. The attribute importance serves as a first indicator which attribute is most influential in affecting respondents' choices. However, these attribute importances only consider the extremes of the partworth utilities and not the intermediate levels. Moreover, the importances can only be interpreted in the context of the selected attributes and levels. Additionally, the attribute importance has to be evaluated in the context of the ability to discriminate between market offerings (Bauer et al. 1996). For example, most ebook readers on the market are 6-in. models. Although the attribute is the second most important based on the range of partworth utilities, it is less managerially relevant since most manufacturers are already offering the most preferred size so that using this attribute level does not help to differentiate from the competitors.

Willingness-to-Pay

The willingness-to-pay (WTP) transformation is based on the idea to analyze how much utility is lost (gained) when the price increases (decreases) and to relate this utility difference to the partworth utility of an attribute level. As a result, the partworth utilities for nonprice attributes can be expressed in monetary terms (Orme 2001).

The WTP calculation requires a vector model for the price attribute, which means that these analyses are only meaningful if the price function is indeed linear. The WTP for level m of attribute n can then be derived by dividing the partworth utility for the specific attribute level by the value of the price vector:

$$WTP_{nm} = \frac{\beta_{nm}}{\beta_p} \tag{11}$$

with,

β_{nm}: partworth utility for level m of attribute n
β_p: utility vector for the price attribute

The estimate for the price vector in the ebook reader example is −0.028, i.e., if price increases by one Euro utility drops by 0.028 units (see Table 6 below). The WTP values for the color attribute can then be calculated as 0.240/−0.028 = €−8.57

Table 6 Estimation results of alternative modeling approaches

Attributes	Partworth model	Vector model for storage and price	Ideal point model for screen size	Interaction effect between screen size and color
Log-likelihood	−2277.8	−2278.3	−2278.3	−2273.3
Storage				
4 GB	−0.389			
8 GB	−0.051			
16 GB	0.440			
(linear)		0.067	0.067	0.067
Screen size				
5 in.	−0.049	−0.050		−0.044
6 in.	0.446	0.446		0.454
7 in.	−0.397	−0.396		−0.410
(linear)			7.854	
(squared)			−0.669	
Color				
Black	−0.002	−0.003	−0.003	−0.015
White	0.240	0.240	0.240	0.255
Silver	−0.238	−0.237	−0.237	−0.240
Price				
€79	0.840			
€99	0.286			
€119	−0.284			
€139	−0.842			
(linear)		−0.028	−0.028	−0.028
No-choice				
	−0.532	−2.965	19.632	−2.984
Screen size × color				
5 in. × black				−0.123
6 in. × black				0.173
7 in. × black				−0.051
5 in. × white				0.018
6 in. × white				−0.164
7 in. × white				0.146

for the color white, €0.07 for black, and €8.50 for silver. The interpretation of these values is that if an ebook reader is not available in, e.g., the preferred color white consumers would accept this drawback only if the price of the reader was, on average, at least €8.57 cheaper. In this case, the negative utility difference of a nonwhite reader is balanced with the positive utility difference of a cheaper price. Vice versa, a consumer would accept paying €8.57 more for a white ebook reader, on average. The least preferred color is silver and consumers would be willing to spend €17.07 for upgrading from a silver ebook reader to a white product. The WTP values can therefore be interpreted directly in terms of consumers' *incremental* willingness to pay for differences in attribute levels. Note, however, that the interpretation needs to consider the differences in signs, i.e., attribute levels with positive utilities have a negative WTP and vice versa.

Market Simulations

The most common ebook readers on the market, e.g., the Amazon Kindle, currently feature 4 GB storage, a 6-in. screen, in the color black for €139. To see how likely it is that consumers buy this product or no ebook reader at all, purchase probabilities can be calculated by applying the MNL function (Eq. 8). These calculations require the specification of a market scenario. A scenario consists of assumptions about the products that are available on the market, i.e., about S, which could include multiple products. In this example, we assume that there are two options, the above-mentioned ebook reader and the no-choice option. On the basis of the aggregate-level estimates, the overall utility of the ebook reader is $V_i = -0.389$ (4 GB) + 0.446 (6 in.) − 0.002 (black) − 0.842 (€139) = −0.787. The utility of the no-choice option is $V_j = -0.532$, i.e., consumers are more likely to buy no ebook reader compared to the one available. The purchase probability for the reader can be calculated by applying Eq. (8):

$$p(i|S) = \frac{\exp(-0.787)}{(\exp(-0.787) + \exp(-0.532))} = 0.437$$

That is, the probability that the sample buys the ebook reader is 43.7%. Market simulations then offer the possibility to see how the market will react if the product configuration is changed. If, e.g., the storage is increased to 8 GB, the overall utility increases to $V_i = -0.449$ and the purchase probability to 0.521. Thus, this modification would be sufficient to make consumers more likely to buy an ebook reader compared to not buying one. Purchase probabilities can be increased further by changing the color to white or reducing the price. These simulations therefore allow detecting promising product modifications. Moreover, a company that wants to enter the market can identify attractive product concepts and assess their effect on purchase probabilities given a specific market scenario that could also consider competitor products. Sophisticated simulation procedures also consider optimal competitive reactions and resulting Nash equilibria (Allenby et al. 2014).

Changing the price in a market simulation, ceteris paribus, allows creating a demand function. In the example above, the purchase probability for the ebook

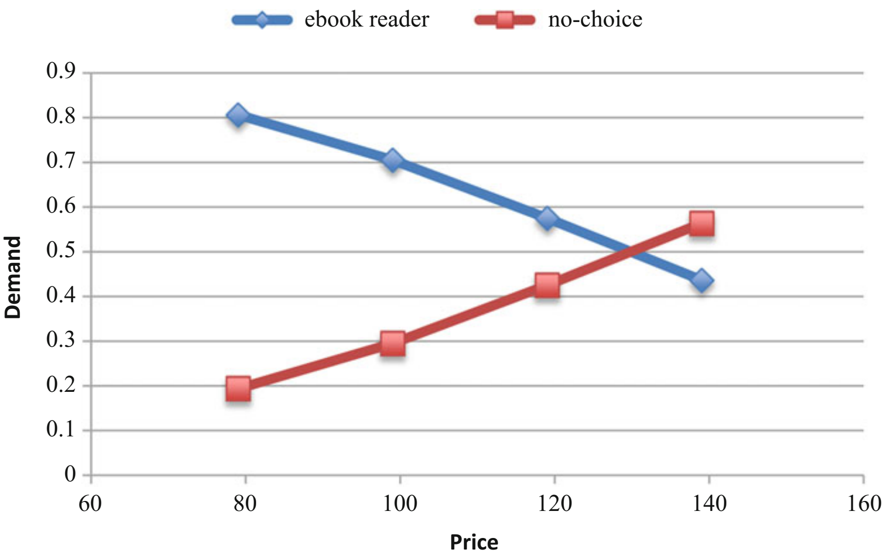


Fig. 8 Demand function for the ebook reader example

reader for €139 is 0.437. Systematically reducing the price increases the probabilities to 0.575 for €119, 0.705 for €99, and 0.806 for €79. The resulting demand function is depicted in Fig. 8. This function can be used to analyze the price elasticity or constitutes an alternative way to calculate WTP. In this example, the price that makes consumers indifferent between choosing the ebook reader and the no-choice option can be taken as the consumers' *absolute* WTP, in this case €130.

The purchase probabilities are frequently interpreted in terms of market shares. Interpreting the predicted probabilities as market shares is ambitious because they have to meet several assumptions (Orme and Johnson 2006). Specifically, probabilities are closer to market shares:

(a) The more the experiment resembles reality, i.e., all attributes and levels that affect buyers need to be accounted for and all competitors are included in the market scenario (assumptions that are not met in this example).
(b) The more the real market environment matches the experiment, i.e., all offers are available, e.g., the products are equally distributed, consumers are aware of the available offers, and there are no switching costs between the offers.

Furthermore, predictions are closer the less consumers' choices are influenced by errors that are introduced by the CBC experiment. It has been shown that incentive alignment is a suitable procedure to accomplish more valid answers so that predictions are closer to market shares (Wlömert and Eggers 2016). Moreover, it is often beneficial to consider heterogeneity among consumers via advanced estimation techniques.

Modeling Alternatives

Table 6 depicts the estimation results of alternative modeling approaches. Next to the partworth model interpreted above, it also shows an approach that uses a vector model for storage and price, i.e., that uses their numeric values instead of its effect codes. It can be seen that the model fit changes only marginally as the difference in log-likelihood is only -0.5, despite using three parameters less. According to a likelihood ratio test this difference is not significant ($p = 0.793$), i.e., this vector model achieves the same fit, while being more parsimonious. The vector model shows that utility increases linearly by 0.067 with every additional GB storage and decreases by -0.028 with every Euro more in purchase price. As the attributes are orthogonal, the other estimates remain largely unaffected. Only the estimate of the no-choice option changes substantially because the numeric values of storage and price are not zero-centered, unlike using effect-coding. This shift does not affect the implications, however.

The third model shown in Table 6 demonstrates the estimation of an ideal-point model for screen size. It requires two parameters, one for the linear effect and one for the squared term. Its model fit is therefore identical to the model in which screen size is represented by a partworth model, which also uses two parameters. The utility for screen size is given by the function $v_{\text{screensize}} = 7.854 * \text{screensize} - 0.669 * \text{screensize}^2$. Accordingly, the ideal point can be calculated as $\partial v / \partial \text{screensize} = 7.854 - 2 * 0.669 * \text{screensize} = 0$, which shows a maximum at 5.87 in.

Finally, the last column of Table 6 adds an interaction effect between the attributes screen size and color. Screen size and color are both represented by two parameters so that 2 * 2 additional parameters are required. Adding these four parameters significantly increases the model fit ($p = 0.039$), i.e., there is an interaction between these two attributes. Accordingly, consumers prefer a black ebook reader in 6 in. and a white version in 7 in.

Advanced Estimation Techniques

The assumption of aggregate-level analyses that consumers are all identical is usually too restrictive. Considering consumer heterogeneity with advanced estimation techniques is therefore beneficial in reducing the error term. Finite mixture (latent class) procedures assume that the sample consists of distinct segments and estimates different utilities for these segments. Continuous mixture (hierarchical Bayes) models are able to estimate individual-level partworth utilities by assuming that the utilities are drawn from a common distribution, e.g., normal distribution. As a result, partworth utilities are generated for each segment or each individual. These values can subsequently be interpreted analogously to the procedures described in sections "Relative Attribute Importance," "Willingness-to-Pay," and "Market Simulations."

Segment-Level Estimation

Segment-level estimation procedures, i.e., latent class estimation, are assuming that a finite number of (homogeneous) segments can represent the heterogeneity of the respondents in the sample. A segment-level perspective is also in line with discovering market segments with distinct preferences that are an attractive target group for a company's market offerings (i.e., following the segmentation, targeting, and positioning approach).

There are two general approaches for segmentation. The first approach determines segments based on socio-demographic data, e.g., separating males and females and estimating aggregate-level preferences for each of these segments. This a priori segmentation, however, is usually not able to detect segments that reflect systematically different preferences towards the attribute levels. The second approach, i.e., the latent class procedure, aims at finding segments that differ in their choice behavior and estimates segment-specific partworth utilities. These segments are latent, i.e., each respondent belongs to the segments with a certain probability (DeSarbo et al. 1995). If a consumer differs in his/her choice behavior from the partworth utilities of the respective segment, this is reflected by a lower probability to belong to this segment (Teichert 2001b).

Before the estimation starts, the researcher needs to define a specific number of segments. In a first step of an iterative-recursive procedure, the segment-specific partworth utilities for the given number of segments are estimated via maximizing the likelihood function. Afterwards, the utility functions are evaluated given the individual respondent's choices in order to allocate the respondents probabilistically to the segments. This results in posterior probabilities of segment membership based on conditional probabilities according to Bayes' rule (DeSarbo et al. 1995). These calculated probabilities form the basis for the iterative process of re-estimating segment-specific utilities. This loop is repeated until only minor changes in the probabilistic allocation of respondents to segments are observed (Sawtooth 2004).

The iterative-recursive process should be repeated for several numbers of segments. The "optimal" number of segments is not determined by the algorithm and has to be based on information criteria, e.g., AIC, BIC, or CAIC (Wedel and Kamakura 2000; Sawtooth 2004). Moreover, a measure of entropy should be inspected, which reflects the accuracy of the segmentation. It is based on the posterior membership probabilities of the respondents. The entropy can exhibit values in the interval [0, 1] and values close to "1" indicate that the segments are well separated, i.e., respondents can be allocated to one of the segments with almost certainty (DeSarbo et al. 1995).

By weighing the segment-level estimates with the membership probability, individual level estimation can be calculated. However, these values lie in the convex hull of the segment-specific utilities so that it is questionable if they can represent individual-level data well (Wedel et al. 1999). Applying the hierarchical Bayes procedures is more appropriate to estimate individual-level preferences (see next chapter).

Table 7 Segment-level estimates

Attributes	Segment 1	Segment 2	Segment 3
Relative segment size	0.592	0.249	0.158
Storage			
4 GB	−0.323	−0.544	−1.195
8 GB	−0.102	0.122	−0.091
16 GB	0.425	0.422	1.286
Screen size			
5 in.	−0.243	0.815	−0.945
6 in.	0.859	0.011	0.108
7 in.	−0.616	−0.826	0.837
Color			
Black	0.302	−0.593	−0.067
White	−0.240	1.246	0.259
Silver	−0.062	−0.653	−0.192
Price			
€79	1.009	0.920	1.423
€99	0.425	0.411	−0.195
€119	−0.318	−0.250	−0.542
€139	−1.116	−1.081	−0.686
No-choice			
	0.118	−2.383	−0.876

Applying the latent class estimation procedure with three segments to the ebook reader case results in a log-likelihood value of −2056.6, i.e., an acceptable McFadden's R^2 of 0.258. The entropy value of 0.948 shows a good separation between the segments. The segment-specific partworth utilities are depicted in Table 7 (not showing standard errors and t-values for better readability).

Based on the membership probabilities, segment 1 is the largest segment with about 60% of the respondents. Segment 2 includes a quarter of the sample and segment 3 follows in size with about 15%. As in the aggregate-level case, the estimates for storage and price show face validity for each segment. Moreover, the segmentation is able to discover segments that prefer smaller screens (segment 2) and larger screens (segment 3). The color white is preferred by segments 2 and 3, however, not by segment 1 that prefers black ebook readers. Finally, segment 1 shows a positive value for the no-choice option, which reflects that this segment is more likely to choose no ebook reader compared to the other segments.

Note that in the aggregate-level analysis 6-in. screens and the color white are preferred by the sample. The conclusion to launch this kind of ebook reader would have been suboptimal as none of the segments prefer this product, i.e., segment 1 prefers 6-in. screens but not the color white, and segment 2 and 3 prefer white but smaller or larger screens.

Individual-Level Estimation

An estimation of individual-level partworth utilities with the MNL model is possible with the hierarchical Bayes (HB) procedure. The idea of the procedure is that the aggregate sample is used to determine the distribution of partworth utilities. The distribution then serves as a basis to draw conditional estimates for each individual given the respondent's choice data. The HB model therefore consists of two coupled layers (Lindley and Smith 1972). The first model layer describes the choice probabilities given the individual partworth utilities, i.e., the MNL model (Eq. 8). The second layer relates the respondents' partworth utilities to each other by assuming a multivariate (normal) distribution of the utilities with unknown mean (Arora et al. 1998).

The model parameters can then be estimated in an iterative process, e.g., with the Metropolis-Hasting algorithm (Chen et al. 2000). Figure 9 depicts the sequence of the HB procedure.

The researcher first needs to specify the type and parameters of the distribution of the utilities. Based on the distribution and the observed choice data, estimates for the individual partworth utilities are drawn in an iterative recursive process. These utilities, in turn, affect the parameters of the distribution, which then serves as a basis to draw a new set of individual-level partworth utilities in a next iteration. This process runs for a large number of iterations, e.g., 20,000, until the parameters converge. Typically, the first set of individual-level utilities draws is discarded as "burn-in" (Sawtooth 2000). The second set of individual-level draws can be used to make inferences about consumer preferences (Allenby et al. 1995).

Figure 10 shows the distribution of individual-level partworth utilities of the ebook reader dataset as boxplots. The mean and median values are plausible and in

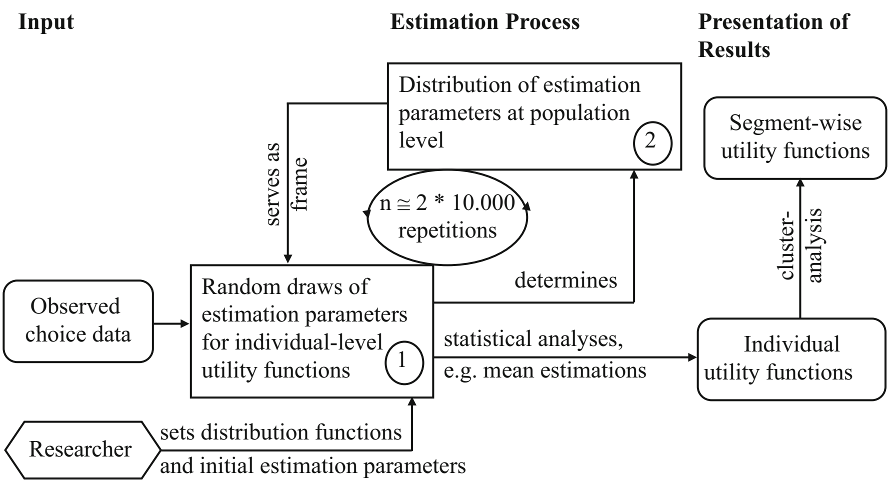
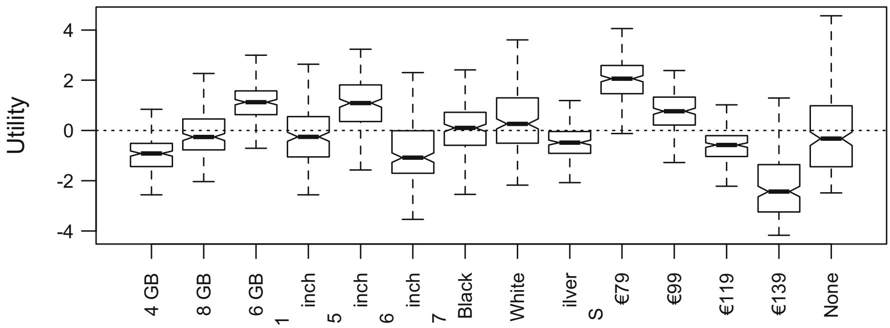


Fig. 9 HB estimation procedures (Teichert 2001b)



	Mean	Standard deviation
Storage		
4 GB	-0.960	0.749
8 GB	-0.173	0.820
16 GB	1.134	0.731
Screen size		
5 inch	-0.187	1.210
6 inch	1.067	0.921
7 inch	-0.880	1.151
Color		
Black	0.026	0.990
White	0.445	1.260
Silver	-0.472	0.668
Price		
€79	2.030	0.804
€99	0.739	0.736
€119	-0.565	0.663
€139	-2.204	1.275
No-choice		
	-0.088	1.593

Fig. 10 Boxplots of partworth utilities and summary statistics for individual-level preferences

line with the aggregate-level model. The distribution and the standard deviation across the respondents' utilities indicate those attributes and attribute levels that exhibit a larger amount of heterogeneous preferences, e.g., screen size, the color white, the highest price, or the no-choice option.

Outlook

Conjoint analysis has emerged from the traditional rating- or ranking-based method in marketing to a general experimental method to study individual's discrete choice behavior with the choice-based conjoint variant. It is therefore not limited to classical applications in marketing, such as new product development, but can be applied to study research questions from related disciplines, e.g., how marketing managers choose their ad campaign, how managers select internationalization options, why consumers engage in or react to social media, etc.

This chapter aims at providing the necessary terminology of conjoint analysis and the requirements to conduct and interpret discrete choice experiments. It also lays the foundation to understand more sophisticated methods and models.

Given the large scope of discrete choice experiments, this attempt is also limited. CBC taps into general theories of how individuals (or groups) choose. These are vast theoretical and empirical grounds, which we cannot cover in detail in this chapter. Understanding CBC models requires not only knowledge of the statistical properties but also understanding behavioral aspects and biases, such as context effects (e.g., compromise, attraction, similarity effects) or trade-off aversion. While knowledge about these aspects is important when running discrete choice experiments, CBC can likewise be used to identify these effects, e.g., by incorporating context effects (Rooderkerk et al. 2011) or by measuring price-quality heuristics (Rao and Sattler 2003).

Although CBC is well developed and documented, many areas are still under research, ranging from, e.g., optimal experimental designs, incentive alignment procedures, to estimation techniques. It will therefore remain an active research area with numerous managerial applications in marketing in the future.

Appendix: R Code

The R code and dataset that correspond to the ebook reader example and estimated models can be found at: http://www.preferencelab.com/data/CBC.R. The estimation uses the mlogit package (Croissant 2012), which needs to be installed first. A less documented version of the R code can be found below (# indicates a comment):

```
# load the library to estimate multinomial choice models.
library(mlogit)

# load (simulated) data about ebook readers
cbc <- read.csv(url("http://www.preferencelab.com/data/
Ebook_Reader.csv"))

# convert data for mlogit
cbc <- mlogit.data(cbc, choice="Selected", shape="long", alt.
var="Alt_id", id.var = "Resp_id")
```

```
### calculate models ###

### partworth model ###
ml1 <- mlogit(Selected ~ Storage_4GB + Storage_8GB +
 Screen.size_5inch + Screen.size_6inch +
 Color_black + Color_white +
 Price_79 + Price_99 + Price_119 +
 None | 0, cbc)
summary(ml1)

# recover reference level estimates (effect-coding)

# Storage_16GB
-(coef(ml1)["Storage_4GB"] + coef(ml1)["Storage_8GB"])

# Screen.size_7inch
-(coef(ml1)["Screen.size_5inch"] + coef(ml1)["Screen.size_6inch"])

# Color_silver
-(coef(ml1)["Color_black"] + coef(ml1)["Color_white"])

# Price_139
-(coef(ml1)["Price_79"] + coef(ml1)["Price_99"] + coef(ml1)
["Price_119"])

# standard errors of the effects are given by the
# square root of the diagonal elements of the
# variance-covariance matrix
covMatrix <- vcov(ml1)
sqrt(diag(covMatrix))

# with effect-coding, the standard error of the reference
# level needs to consider the off-diagonal elements of the
# corresponding attribute levels

# Std. Error Storage_16GB
sqrt(sum(covMatrix[1:2, 1:2]))

# Std. Error Screen.size_7inch
sqrt(sum(covMatrix[3:4, 3:4]))

# Std. Error Color_silver
sqrt(sum(covMatrix[5:6, 5:6]))

# Std. Error Price_139
sqrt(sum(covMatrix[7:9, 7:9]))
```

```
### Vector model ###
# Storage and Price follow a linear trend. Replacing
# parameters leads to a more parsimonious model.

ml2 <- mlogit(Selected ~ Storage +
 Screen.size_5inch + Screen.size_6inch +
 Color_black + Color_white +
 Price +
 None | 0, cbc)
summary(ml2)

# likelihood ratio test
lrtest(ml2, ml1)

# incremental willingness-to-pay for storage
coef(ml2)["Storage"]/coef(ml2)["Price"]

# WTP to upgrade from a black to a white ebook reader
(coef(ml2)["Color_white"] - coef(ml2)["Color_black"])/coef(ml2)
["Price"]

### Vector model for screen size has sig. worse fit ###
ml3 <- mlogit(Selected ~ Storage + Screen.size + Color_black +
Color_white + Price + None | 0, cbc)
summary(ml3)

lrtest(ml3, ml2)

### Testing an ideal point model for screen size ###
ml4 <- mlogit(Selected ~ Storage +
 Screen.size + I(Screen.size**2) +
 Color_black + Color_white +
 Price +
 None | 0, cbc)
summary(ml4)

# same model fit because no differences in df
lrtest(ml4, ml2)

### Adding interactions between screen size and color ###
ml5 <- mlogit(Selected ~ Storage +
 Screen.size_5inch + Screen.size_6inch +
 Color_black + Color_white +
 Price +
 Screen.size_5inch * Color_black +
 Screen.size_6inch * Color_black +
```

```
 Screen.size_5inch * Color_white +
 Screen.size_6inch * Color_white +
 None| 0, cbc)
summary(ml5)

# likelihood ratio test
lrtest(ml2, ml5)
```

References

Addelman, S. (1962). Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics, 4*(1), 21–46.

Allenby, G. M., Arora, N., & Ginter, J. L. (1995). Incorporating prior knowledge into the analysis of conjoint studies. *Journal of Marketing Research, 32*(2), 152–162.

Allenby, G. M., Brazell, J. D., Howell, J. R., & Rossi, P. E. (2014). Economic valuation of product features. *Quantitative Marketing and Economics, 12*(4), 421–456.

American Marketing Association. (2015). American Marketing Association AMA. https://www.ama.org/resources/Pages/Dictionary.aspx. Accessed 15 Nov 2015.

Arora, N., Allenby, G. M., & Ginter, J. L. (1998). A hierarchical Bayes model of primary and secondary demand. *Marketing Science, 17*(1), 29–44.

Batsell, R. R., & Louviere, J. J. (1991). Experimental analysis of choice. *Marketing Letters, 2*(3), 199–214.

Bauer, H., Herrmann, A., & Homberg, F. (1996). *Analyse der Kundenwünsche zur Gestaltung eines Gebrauchsgutes mit Hilfe der Conjoint Analyse.* Universität Mannheim, Lehrstuhl für ABWL und Marketing II, Working Paper Nr. 110.

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9*(3), 226–232.

Brazell, J. D., Diener, C. G., Karniouchina, E., Moore, W. L., Séverin, V., & Uldry, P.-F. (2006). The no-choice option and dual response choice designs. *Marketing Letters, 17*(4), 255–268.

Burmester, A., Eggers, F., Clement, M., & Prostka, T. (2016). Accepting or fighting unlicensed usage – Can firms reduce unlicensed usage by optimizing their timing and pricing strategies? *International Journal of Research in Marketing, 33*(2), 434–356.

Chakraborty, G., Ball, D., Gaeth, G. J., & Jun, S. (2002). The ability of ratings and choice conjoint to predict market shares – A Monte Carlo simulation. *Journal of Business Research, 55*(3), 237–249.

Chen, M.-H., Shao, Q.-M., & Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian computation*. New York: Springer Series in Statistics.

Croissant, Y. (2012). Estimation of multinomial logit models in R: The mlogit packages. *R package version 0.2-2*. http://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf.

De Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in the health economics: A review of the literature. *Health Economics, 21*(2), 145–172.

DeSarbo, W. S., Ramaswamy, V., & Cohen, S. (1995). Market segmentation with choice-based conjoint analysis. *Marketing Letters, 6*(2), 137–147.

Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research, 44*(2), 214–223.

Ding, M., Grewal, R., & Liechty, J. (2005). Incentive-aligned conjoint analysis. *Journal of Marketing Research, 42*(2), 67–82.

Ding, M., Park, Y.-H., & Bradlow, E. T. (2009). Barter markets for conjoint analysis. *Management Science, 55*(6), 1003–1017.

Dong, S., Ding, M., & Huber, J. (2010). A simple mechanism to incentive-align conjoint experiments. *International Journal of Research in Marketing, 27*(1), 25–32.
Eggers, F., & Sattler, H. (2009). Hybrid individualized two-level choice-based conjoint (HIT-CBC): A new method for measuring preference structures with many attribute levels. *International Journal of Research in Marketing, 26*(2), 108–118.
Eggers, F., Hauser J. R., & Selove, M. (2016). The effects of incentive alignment, realistic images, video instructions, and ceteris paribus instructions on willingness to pay and price equilibria. *Proceedings of the Sawtooth Software conference*, 1–18 September.
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.
Frischknecht, B., Eckert, C., Geweke, J., & Louviere, J. J. (2014). A simple method for estimating preference parameters for individuals. *International Journal of Research in Marketing, 31*(1), 35–48.
Gensler, S., Hinz, O., Skiera, B., & Theysohn, S. (2012). Willingness-to-pay estimation with choice-based conjoint analysis: Addressing extreme response behavior with individually adapted designs. *European Journal of Operational Research, 219*(2), 368–378.
Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research, 5*, 103–123.
Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing, 54*, 3–19.
Haaijer, R., & Wedel, M. (2003). Conjoint experiments. general characteristics and alternative model specifications. In A. Gustafsson, A. Herrmann, & F. Huber (Eds.), *Conjoint measurement: Methods and applications* (3rd ed., pp. 371–412). Berlin: Springer.
Haaijer, R., Wedel, M., Vriens, M., & Wansbek, T. (1998). Utility covariances and context effects in conjoint MNP models. *Marketing Science, 17*(3), 236–252.
Haaijer, R., Kamakura, W. A., & Wedel, M. (2001). The "no-choice" alternative to conjoint choice experiments. *International Journal of Market Research, 43*(1), 93–106.
Hartmann, A. (2004). *Kaufentscheidungsprognose auf Basis von Befragungen. Modelle, Verfahren und Beurteilungskriterien*. Wiesbaden: Gabler.
Hensher, D. A. (1994). Stated preference analysis of travel choices: The state of practice. *Transportation, 21*(2), 107–133.
Hensher, D. A., & Johnson, L. W. (1981). *Applied discrete choice modelling*. New York: Wiley.
Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research, 33*(3), 307–317.
Johnson, R. M. (1987). Adaptive conjoint analysis. In *Sawtooth software conference proceedings*. Ketchum: Sawtooth Software.
Johnson, R. M., & Orme, B. K. (1996). *How many questions should you ask in choice-based conjoint studies?* (Sawtooth software research paper series). Sequim: Sawtooth Software.
Kraus, S., Ambos, T. C., Eggers, F., & Cesinger, B. (2015). Distance and perceptions of risk in internationalization decisions. *Journal of Business Research, 68*(7), 1501–1505.
Lindley, D. V., & Smith, A. F. (1972). Bayes estimates for the linear models. *Journal of the Royal Statistical Society, Series B, 34*(1), 1–41.
Louviere, J. J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments. An approach based on aggregated data. *Journal of Marketing Research, 20*(4), 350–367.
Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods. Analysis and application*. Cambridge: Cambridge University Press.
Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods, and applications*. Cambridge: Cambridge University Press.
Lusk, J. L., & Schroeder, T. C. (2004). Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics, 86*(2), 467–482.
McFadden, D. (1981). Econometric models of probabilistic choice. In C. Manski & D. McFadden (Eds.), *Structural analysis of discrete data* (pp. 198–272). Cambridge: MIT-Press.

Meissner, M. Oppewal, H., & Huber, J. (2016). How many options? Behavioral responses to two versus five alternatives per choice. *Proceedings of the Sawtooth Software conference*, 1–18 September.

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.

Moore, W. L. (2004). A cross-validity comparison of rating-based and choice-based conjoint analysis models. *International Journal of Research in Marketing, 21*(3), 299–312.

Moore, W. L., Gray-Lee, J., & Louviere, J. J. (1998). A cross-validity comparison of conjoint analysis and choice models at different levels of aggregation. *Marketing Letters, 9*(2), 195–207.

Orme, B. (2001). *Assessing the monetary value of attribute levels with conjoint analysis: Warnings and suggestions* (Sawtooth software research paper series). Sequim: Sawtooth Software.

Orme, B. (2002). *Formulating attributes and levels in conjoint analysis* (Sawtooth software research paper series). Sequim: Sawtooth Software.

Orme, B. K. (2016). *Results of the 2017 Sawtooth Software User Survey*. https://www.sawtoothsoftware.com/about-us/news-and-events/news/1693-results-of-2016-sawtooth-software-user-survey.

Orme, B., & Johnson, R.M. (2006). *External effect adjustments in conjoint analysis* (Sawtooth software research paper series). Sequim: Sawtooth Software.

Page, A. L., & Rosenbaum, H. F. (1992). Developing an effective concept testing program for consumer durables. *Journal of Product Innovation Management, 9*, 267–277.

Park, Y.-H., Ding, M., & Rao, V. R. (2008). Eliciting preference for complex products: A web-based upgrading method. *Journal of Marketing Research, 45*(5), 562–574.

Rao, V. R., & Sattler, H. (2003). Measurement of price effects with conjoint analysis: Separating informational and allocative effects of price. In *Conjoint Measurement* (pp. 47–66). Berlin/Heidelberg: Springer.

Rooderkerk, R. P., Van Heerde, H. J., & Bijmolt, T. H. (2011). Incorporating context effects into a choice model. *Journal of Marketing Research, 48*(4), 767–780.

Sattler, H. (2005). Markenbewertung: State-of-the-Art. *Zeitschrift für Betriebswirtschaft*, *2*, 33–57.

Sattler, H. (2006). Methoden zur Messung von Präferenzen für Innovationen. *Zeitschrift für Betriebswirtschaftliche Forschung, 54*(6), 154–176.

Sattler, H., Hartmann, A., & Kröger, S. (2004). Number of tasks in choice-based conjoint analysis. *Conference proceedings of the 33rd EMAC conference*. Murcia.

Sawtooth (1999). *The choice-based conjoint (CBC) technical paper* (Sawtooth software technical paper series). Sequim: Sawtooth Software.

Sawtooth. (2000). *The CBC/HB system for hierarchical Bayes estimation version 4.0* (Sawtooth software technical paper series). Sequim: Sawtooth Software.

Sawtooth. (2004). *The CBC latent class technical paper (version 3)* (Sawtooth software technical paper series). Sequim: Sawtooth Software.

Sawtooth. (2013). *The MaxDiff system – Technical paper* (Sawtooth software technical paper series). Orem: Sawtooth Software.

Sawtooth. (2014). *ACBC – Technical paper* (Sawtooth software technical paper series). Orem: Sawtooth Software.

Schlereth, C., & Skiera, B. (2016). Two new features in discrete choice experiments to improve willingness-to-pay estimation that result in SDR and SADR: Separated (adaptive) dual response. *Management Science, 63*(3), 829–842.

Shocker, A. D., & Srinivasan, V. (1973). Linear programming techniques for multidimensional analysis of preference. *Psychometrika*, 337–369.

Sloan, N. J. A. (2015). A library of orthogonal arrays. http://neilsloane.com/oadir/. Accessed 15 Nov 2015.

Srinivasan, V., & Park, C. S. (1997). Surprising robustness of the self-explicated approach to customer preference structure measurement. *Journal of Marketing Research, 34*(2), 286–291.

Teichert, T. (2001a). *Nutzenschätzung in Conjoint-Analysen: Theoretische Fundierung und empirische Aussagekraft*. Wiesbaden: Springer.

Teichert, T. (2001b). Nutzenermittlung in wahlbasierten Conjoint-Analysen. Ein Vergleich zwischen Latent-Class- und hierarchischem Bayes-Verfahren. *Zeitschrift für Betriebswirtschaftliche Forschung, 53*(8), 798–822.

Toubia, O., Simester, D. I., Hauser, J. R., & Dahan, E. (2003). Fast polyhedral adaptive conjoint estimation. *Marketing Science, 22*(3), 273–303.

Toubia, O., Hauser, J. R., & Simester, D. I. (2004). Polyhedral methods for adaptive choice-based conjoint analysis. *Journal of Marketing Research, 41*, 116–131.

Toubia, O., Hauser, J., & Garcia, R. (2007). Probabilistic polyhedral methods for adaptive choice-based conjoint analysis: Theory and application. *Marketing Science, 26*(5), 596–610.

Toubia, O., de Jong, M. G., Stieger, D., & Füller, J. (2012). Measuring consumer preferences using conjoint poker. *Marketing Science, 31*(1), 138–156.

Train, K. (2009). *Discrete choice models with simulation* (2nd ed.). Cambridge: Cambridge University Press.

Urban, G. L., & Hauser, J. R. (1993). *Design and marketing of new products* (2nd ed.). Englewood Cliffs: Prentice Hall.

Urban, G. L., Weinberg, B. D., & Hauser, J. R. (1996). Premarket forecasting of really-new products. *Journal of Marketing, 60*(1), 47–60.

Verlegh, P. W. J., Schifferstein, H. N. J., & Wittink, D. R. (2002). Range and number-of-levels in derived and stated measures of attribute importance. *Marketing Letters, 13*(1), 41–52.

Voeth, M. (1999). 25 Jahre conjointanalytische Forschung in Deutschland. *Zeitschrift für Betriebswirtschaft*, Ergänzungsheft 2, 153–176.

Vriens, M., Oppewal, H., & Wedel, M. (1998). Rating-based versus choice-based latent class conjoint models – An empirical comparison. *Journal of the Market Research Society, 40*(3), 237–248.

Walker, J., & Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences, 43*(3), 303–343.

Wedel, M., & Kamakura, W. A. (2000). *Market segmentation. conceptual and methodological foundations* (2nd ed.). Boston: Springer.

Wedel, M., Kamakura, W. A., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R. M., Lenk, P., Neslin, S., & Poulsen, C. S. (1999). Discrete and continuous representations of unobserved heterogeneity in choice modeling. *Marketing Letters, 10*(3), 219–232.

Wertenbroch, K., & Skiera, B. (2002). Measuring consumers' willingness to pay at the point of purchase. *Journal of Marketing Research, 39*(2), 228–241.

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*, 41–52.

Wlömert, N., & Eggers, F. (2016). Predicting new service adoption with conjoint analysis: External validity of BDM-based incentive-aligned and dual-response choice designs. *Marketing Letters, 27*(1), 195–210.

Zeithammer, R., & Lenk, P. (2009). Statistical benefits of choices from subsets. *Journal of Marketing Research, 46*(6), 816–831.