Huertas-García, Rubén; Nuñez Carballosa, Ana; Miravitlles, Paloma

Article

Statistical and cognitive optimization of experimental designs in conjoint analysis

European Journal of Management and Business Economics (EJM&BE)

Provided in Cooperation with:
European Academy of Management and Business Economics (AEDEM), Vigo (Pontevedra)

Suggested Citation: Huertas-García, Rubén; Nuñez Carballosa, Ana; Miravitlles, Paloma (2016) : Statistical and cognitive optimization of experimental designs in conjoint analysis, European Journal of Management and Business Economics (EJM&BE), ISSN 2444-8451, Elsevier, Amsterdam, Vol. 25, Iss. 3, pp. 142-149, https://doi.org/10.1016/j.redee.2015.10.001

European Journal of Management and Business Economics 25 (2016) 142–149

European Journal of Management and Business Economics

www.elsevier.es/ejmbe

Article

Statistical and cognitive optimization of experimental designs in conjoint analysis

Rubén Huertas-García[a,*], Ana Nuñez-Carballosa[b], Paloma Miravitlles[b]

[a] Comercialización e Investigación de Mercados, Departamento de Economía y Organización de Empresas, Facultad de Economía y Empresa, Universidad de Barcelona, Barcelona, Spain
[b] Organización de Empresas, Departamento de Economía y Organización de Empresas, Facultad de Economía y Empresa, Universidad de Barcelona, Barcelona, Spain

ARTICLE INFO

Article history:
Received 2 April 2015
Accepted 3 October 2015
Available online 17 December 2015

Keywords:
Optimal experimental design
Cognitive ability
Conjoint analysis
Preference measurement
Dynamic algorithm

ABSTRACT

Conjoint analysis has become the most used technique for measuring preferences for new products to be launched in the market. Experimental design models are key elements for its use in market research. Such models involve a matrix in which attributes and levels are combined, making product concepts that respondents then evaluate.

Experimental design has emerged as a key element in conjoint analysis' success because its application generates statistical and reliability implications for part-worth factor estimations and for the type of heuristics followed by respondents.

This paper proposes a conceptualization of both statistical and cognitive efficiency criteria for experimental designs. A review of the most used statistical optimization criteria is presented, as well as a methodology for optimizing cognitive efficiency. Finally, we suggest a dynamic algorithm for optimizing the objective function in a sequential manner.

Introduction

Conjoint analysis (CA) has been one of the most successful market research techniques in the last 40 years and one of the reasons that explains this is the continuous interaction between academic theory and professional implementation (Bradlow, 2005; Green, Krieger, & Wind, 2001). Its success is not only due to the technical upgrade linked to its application in studies on consumer preferences among products with multiple attributes, but also to its versatility, which has permitted its application in diverse fields such as marketing, transport management, financial services (Green et al., 2001), and even oncological studies (Beusterien et al., 2014).

Although the statistical foundations of experimental design were developed throughout the 1920s by Ronald Fisher, it was not until the 1970s that researchers began to use it in psychology and, soon after, in consumer behaviour and market research studies (Gustafsson, Herrmann, & Huber, 2007). However, despite the numerous investigations in which CA has been used (e.g. when someone introduces the term 'conjoint analysis' in the 'Web of Science' there are more than 3000 references) and the amount of methodological innovations that have been proposed, such as adaptive conjoint analysis (Johnson, 1987) or polyhedral designs (Toubia, Simester, Hauser, & Dahan, 2003), there are some issues in conjoint experimental designs that remain unresolved. These include the criteria for choosing the number of attributes and the reasons for using an experimental design and no other (Bradlow, 2005).[1]

This paper attempts to deepen the analysis of these two problems from a dual perspective: firstly, through the consideration of an efficient statistical solution and, secondly, taking into account the cognitive burden of these designs and its effect on responses. The literature proposes different criteria and ways to qualify an experimental design as efficient and each criterion determines the type of design recommended and the type of analysis to be used on the data gathered. Although different methods have different structures and philosophies, they nevertheless share some common elements: firstly, they all use the statistical design of experiments to develop the experimental design and, secondly, they all consider a compensatory model of consumer behaviour that, moreover, considers this behaviour as independent of the experimental design (Johnson & Meyer, 1984). Although the first common element

* Corresponding author.
E-mail address: rhuertas@ub.edu (R. Huertas-García).

[1] Although Bradlow (2005) points out nine problems in the use of conjoint analysis, in this study we have focused our analysis on two of them.

http://dx.doi.org/10.1016/j.redee.2015.10.001

in experimental design appears to remain valid, the second has received numerous criticisms (Olshavsky, 1979; Payne, Bettman, & Johnson, 1988). In this paper, we review the most common criteria used to determine experimental design efficiency, as well as their advantages and disadvantages, and, in addition, we propose a methodology to estimate the cognitive efficiency of experimental designs.

The study is arranged in the following sections. Firstly, we present some of the most important concepts of experimental design in CA. Secondly, we describe the most common methods for determining statistical efficiency in experimental designs. We then propose a system for evaluating the cognitive efficiency of experimental designs, which is illustrated with an example. Finally, some conclusions and possibilities for further research are presented.

Experimental design in conjoint analysis

Experimental design is a numerical arrangement that combines attributes and levels to form stimuli (concepts of products or services), which later, in field work, must be evaluated by respondents. There are several ways of organizing stimuli, with full factorial designs being the easiest because they simply require presenting all possible level and factor combinations.

Experimental design is represented by an $n \times k$ matrix, where n are rows indicating the profiles generated and k are columns indicating the level variations of the attributes. If you use combinatory laws in the design, it is called a statistical design (Box, Hunter, & Hunter, 2005). However, statistical designs pose a major problem since as new attributes are added to the experimental design (even if they have only 2 or 3 levels), the number of combinations grows exponentially. For example, a statistical design in which there are 4 attributes with 3 levels each generates an experimental design of 81 profiles ($3^4 = 81$). Such a number of alternatives can overwhelm the cognitive ability of any interviewee (Green & Srinivasan, 1990). For this reason, researchers use fractional factorial designs, which are simply portions of a full factorial design. This reduces the number of alternatives that each interviewee has to assess, but researchers must pay the price of sacrificing the ability to estimate certain interactions. A price which, in some cases, could be very high, as for example in the research of new products associated with the sensory perceptions of respondents (haircare and cosmetic products) where two-factor interactions are very important (Green & Srinivasan, 1990).

Therefore, it has been shown that CA works quite well when a moderate number of attributes is considered for each profile within the experimental design, for example, less than 8 (Bradlow, 2005). But when the number of attributes is high, for example, more than 15, respondents' cognitive capacity is overwhelmed and they use non-compensatory heuristics to choose profiles (Hauser, 2014; Johnson & Meyer, 1984). However, many technological products, such as laptops, digital cameras, and even cars or tourist services, use a large amount of attributes in their advertising to differentiate themselves from competitors (Bradlow, 2005; Netzer & Srinivasan, 2011).

In order to address this problem, several solutions have been proposed, in some cases hybrid solutions using a combination of techniques in a single study. For example, this is the case in the use of adaptive conjoint analysis (Johnson, 1987) or, more recently, adaptive self-explicated approaches (Netzer & Srinivasan, 2011). These methods follow two stages: in the first stage, respondents must complete a self-explicated survey (compositional study) where different attributes and levels are measured with the aim of reducing their number to the most relevant and, in the second stage, there is a conjoint experiment (decompositional study) to estimate the part-worth of the factors.

Now, for any of these experimental designs to generate valid and reliable data they must follow two basic principles: orthogonality and balance (Chrzan & Orme, 2000). An experimental design is orthogonal if all main effects and interactions can be calculated as independent variables (i.e. without correlation between them), and it is balanced when each level of each factor is repeated the same number of times in the overall experiment. When an experimental design is orthogonal and balanced at the same time, it is said that it is optimal (Kuhfeld, Tobias, & Garrat, 1994). However, when an experimental design is formed by factors with different ranges of levels (i.e. factors have 2, 3, 4, or 5 levels combined in the same design), it is very difficult to make fractions that are both orthogonal and balanced. Therefore, there is no other option than to use quasi-orthogonal designs (Kuhfeld, 1997) and, in this case, it is necessary to have some measure of their efficiency.

Criteria of efficiency

In recent years, the literature has shown great concern around finding and developing efficient statistical designs (Vermeulen, Goos, & Vandebroek, 2008). A design is considered efficient when it gets the most and best information possible with the least number of interviewees and with the shortest time dedicated to field work (Louviere, Islam, Wasi, Street, & Burgess, 2008). However, knowing the researcher's objectives and assumptions is a prerequisite for trying to achieve efficiency in an experimental design. That is, whether the researcher considers interactions between factors or not, and the extent of these interactions, or, if they take into account the type of protocol used to gather information; i.e. if she or he uses verbal descriptions, images, or combinations of both. This means that the same experimental design can be efficient with regard to one criterion but not another. But the literature that discusses design efficiency only considers statistical efficiency (Kuhfeld et al., 1994; Vermeulen et al., 2008).

Kiefer (1959) was the first to propose the statistical efficiency concept as an instrument to compare and evaluate experimental designs and he described it using the so-called 'theory of optimal design'. The idea was to try to gather all of a statistical design's goodness-of-fit in a figure, i.e. to transform design variance and covariance into a number, facilitating the comparison of various experimental designs and choosing the design that minimizes or maximizes that number. In addition, there are some criteria that depend on the objectives pursued by each researcher. For example, the researcher might only be interested in the estimation of main factors or she might also be interested in two-factor interactions, and in each case she must use different criteria to optimize the design.

The different criteria are represented by letters following alphabetical order, among the most cited are criteria A and D (Kuhfeld et al., 1994; Vermeulen et al., 2008) and, moreover, they are also the most used by the software packages (Kuhfeld, 1997; Myers, Montgomery, & Anderson-Cook, 2009). However, although the enormous virtues offered by computer-generated designs are recognized, especially because they are easy to get, there are some authors who criticize how they are used indiscriminately and without taking into account the criteria behind the design or, therefore, the type of adjustment that can be achieved with the design obtained (Myers et al., 2009). Thus, knowing the basic characteristics of these designs can help researchers to select one criterion or another depending on the main purpose of the research and bearing in mind the limitations that each criterion entails.

The optimization criteria are based on the moment matrix. To illustrate its operation, we assume that a conjoint experiment was carried out, it was coded following a vectorial form, and in it respondents rated each profile from 1 to 10. In addition, it was assumed

that these scores reflected respondents' preferences in a linear model of part-worth attributes (Johnson & Meyer, 1984). In this case, the easiest way to estimate the part-worth of factors is to adjust them by linear regression, where the dependent variable μ_j gathers the scores assigned to each profile for $j = \{1, \ldots, n\}$:

$$\mu_j = \beta_0 + \sum_{i=1}^{q} \beta_i x_{ij} + \varepsilon_j, \tag{1}$$

where x_{ij} are experimental design combinations, β_i are part-worth estimates representing factor slopes, and ε_j is the error. In matrix notation, it can be expressed as $\boldsymbol{\mu} = \alpha + \beta\mathbf{X} + \mathbf{e}$, $\mathbf{X}$ being the experimental design matrix coded in vector form (−1, 0, 1). Efficiency criteria are based on the so-called moment matrix $\mathbf{M}$, which is defined as:

$$\mathbf{M} = \frac{\mathbf{X}'\mathbf{X}}{N}, \tag{2}$$

where $\mathbf{X}'\mathbf{X}$ is the product of the transposed matrix multiplied by the normal matrix from the experimental design and N is the number of profiles.

The A-optimal criterion only considers variance as a relevant element for defining efficiency. This criterion consists of calculating the sum of variances of the model parameters and it is the same as the sum of diagonal parameters, or the trace, of $(\mathbf{X}'\mathbf{X})^{-1}$. This means that the A-optimal criterion prioritizes main factor estimations without taking into account interactions that are reflected in covariance. Therefore, the optimization criterion consists of choosing the experimental design that minimizes the trace of the inverse moment matrix and it will be defined by:

$$\min_{\zeta} tra[\mathbf{M}(\zeta)]^{-1}, \tag{3}$$

where *tra* represents the trace of the moment matrix, i.e. the sum of coefficient variance, and ζ represents the design used.

The other most used criterion is the D-optimal criterion (Toubia & Hauser, 2007; Vermeulen et al., 2008). This also seeks to find the experimental design that improves parameter estimation, but in this case, besides the main factors, it also includes interactions. This criterion requires calculating the determinant of the moment matrix:

$$|\mathbf{M}| = \frac{|\mathbf{X}'\mathbf{X}|}{N^p}, \tag{4}$$

where p is the number of model parameters (which can be main factors, two-factor interactions, three-factor interactions, etc.) and N is the number of profiles analysed. If we assume that errors follow a normal distribution, independent and with constant variance, the determinant of the moment matrix $(\mathbf{X}'\mathbf{X})$ is inversely proportional to the confidence intervals of coefficients in the regression model, and this reflects how well these coefficients are estimated. In this case, a small value in $|\mathbf{X}'\mathbf{X}|$ implies a poor estimate of β (Myers et al., 2009). Thus, a design is D-optimal if it minimizes the determinant of the moment matrix (4), which is the same as the inverse of the determinant of the variance and covariance matrix:

$$(\mathbf{X}'\mathbf{X}) = \frac{1}{|(\mathbf{X}'\mathbf{X})^{-1}|}, \tag{5}$$

Therefore, the aim of the D-optimal criterion is to establish which experimental design minimizes both variance and covariance. Furthermore, due to the fact that this criterion uses the power p (number of parameters to be estimated) in the calculation of the determinant of the moment matrix, it allows for use of the D-optimal efficiency criterion to compare experimental designs of different sizes.

In short, whether we use A-optimal or D-optimal criteria, if we have a balanced and orthogonal experimental design its efficiency will be optimal for both, and when we analyse quasi-orthogonal designs they will be more efficient as they tend towards balance and orthogonality. These criteria measure the design's goodness-of-fit in relation to a hypothetical orthogonal design, which it is not possible to use for multiple reasons. However, these measures should not be considered as absolute measures of design efficiency, but as relative tools to compare one design with another. An efficient design must fulfil the following requirements (Kuhfeld, 1997):

- A design is orthogonal and balanced when the moment matrix $\mathbf{M}$ is diagonal.
- A design is orthogonal when the sub-matrix $\mathbf{M}$, excluding the row and column for the intercept variable, that is to say α, is diagonal.
- A design is balanced when all off-diagonal elements, all row and column interceptions, are zero.
- A design's efficiency increases as diagonal absolute values become smaller, that is to say, as N^p becomes larger.

If a full factorial design is orthogonal and balanced (i.e. the moment matrix $\mathbf{M}$ is diagonal), it is possible to use the A-optimal or D-optimal criteria indifferently. However, if the design is fractional factorial will be some off-diagonal values, it must be the researcher who decides if she is interested only in main factor estimations, thus using the A-optimal criterion, or if she is also interested in two-factor interactions, in which case the best approach would be to use the D-optimal criterion.

The goal pursued by Kiefer (1959) of reducing the statistical goodness-of-fit of an experimental design to a single number has been criticized as overly ambitious and, therefore, these criteria should be considered as one of many other relevant aspects, but not as a definitive measure (Myers et al., 2009). On the other hand, CA literature has also criticized that accepting these criteria as valid involves the assumption that respondents value profiles using an algebraic utility function, where the assessment of each profile is based exclusively on the observed attributes, and that their assessment is independent of the number of profiles to be evaluated (Johnson & Meyer, 1984). In the same vein, but in choice-based conjoint analysis, Louviere (2001) warns that the use of increasingly efficient designs can generate unintended consequences in respondents' answers due to their cognitive limitations, and they may therefore use heuristics other than the compensatory model during their selection process (Hauser, 2014).

In short, a statistically efficient experimental design does not guarantee the responses' reliability or good predictive capability, since these will depend on the effort and sincerity of interviewees' answers. CA is also used by academic researchers and market researchers who highly value the predictive capability of the models obtained. On the other hand, if the data gathered are not accurate, they can lead to misdiagnoses and, thus, to erroneous advice about management policies (Salisbury & Feinberg, 2010). Therefore, when an experimental design is developed we should not only take into account its statistical efficiency, but should also incorporate some criteria to measure its cognitive efficiency.

Determinants of cognitive effort in experimental designs

A common issue in CA literature is understanding how experimental designs determine the cognitive effort that respondents have to make and how their effort affects response reliability and its predictive capability. In particular, from a cognitive perspective, what would be the optimum size of a choice set?

Different theoretical approaches have proposed different behavioural models depending on their theoretical axioms. For

example, economic theory considers that if a consumer has a large number of profiles to choose from, this large choice set improves her likelihood of finding a profile with level-factor combinations that better fit her preferences (Lancaster, 1990). However, this rational model has been criticized from the perspective of other theoretical frameworks. Johnson and Meyer (1984) advocate the opposite: that it is better to maintain choice sets with a small number of profiles because a larger amount would have negative statistical and cognitive effects (see also DeShazo & Fermo, 2002; Louviere et al., 2008). From the statistical point of view, having many attributes (whether profiles or levels) in a choice set reduces the amount of data available for each attribute, undermining their weight and the accuracy of estimates (DeShazo & Fermo, 2002). In fact, as noted above, Bradlow (2005) considers that CA works relatively well with around 8 attributes per profile and that problems begin to be generated from 15 attributes and beyond.

Regarding cognitive effects, information processing theory considers that people have limited rational capacity and that when they face a decision-making process, the greater the number of alternatives, the greater the difficulty in processing, calculating, and storing all the information generated throughout the process (Simon, 1990). According to Johnson and Meyer (1984), the complexity of a choice is determined by the number of cognitive steps that each respondent has to take to carry out an evaluation and make the appropriate decision, comparing all of the alternatives. For example, if a consumer compares two photographic cameras defined by two attributes (one has 'the capacity to take 300 pictures' and one is 'pocket-size') the respondent has to make at least three assessments or 'cognitive steps' in Johnson and Meyer's terms: (1) assess if 'the capacity to take 300 pictures' is a good attribute; (2) evaluate if 'pocket-size' is also a good attribute; and (3) compare the two attributes and decide which one is best; as the number of attributes increases, the number of assessments and 'cognitive steps' also increases. Because the respondent comes to have difficulties in her evaluation of attributes one by one, she begins to use an alternative criterion of simplified heuristics such as the lexicographic heuristic or elimination by aspect (Hauser, 2014; Payne, 1982). As a result of using inconsistent responses and of not using the compensatory model, the linear model estimates will generate a poor fit (Payne et al., 1988).

A proposal to optimize cognitive effort in experimental designs

In this study, we make use of the cognitive complexity concept proposed by Johnson and Meyer (1984) and we also use the information acquisition cost function proposed by Grether and Wilde (1984). Both criteria where considered for optimizing an experimental design taking into account all steps taken by a respondent in making the appropriate decision. To achieve this optimum it is necessary to determine the number of profiles, factors, and levels that minimize the cost of the whole assessment process of a choice set.

One useful criterion for estimating the cost of cognitive effort has been the time spent on making a decision (Dellaert, Donkers, & Soest, 2012; Grether & Wilde, 1984; Johnson & Meyer, 1984). The time spent on a decision remains a recurring issue both in consumer behaviour studies (East, Wright, & Vanhuele, 2013) and in experimental design studies (Louviere et al., 2008). For example, in a classic experiment, East (1973) showed that respondents needed more time to choose between two alternatives than between three when the choice task was difficult and less time when the choice task was easier. This means that choosing between a larger number of profiles does not always require more time than choosing between a smaller number. However, time is a variable that is very closely related to effort (Grether & Wilde, 1984).

Besides, it is also important to consider the time taken to complete the questionnaire, because this can have negative consequences on respondents' behaviour: the longer the time required to do the interview, the lower the willingness to participate and the greater the chance of it being left unfinished (Netzer & Srinivasan, 2011; Scholz, Meissnery, & Decker, 2010). It is clear that as the experimental design's complexity increases (the choice set changes from 2 to 4 and then to 6 cameras), so does its duration (Grether & Wilde, 1984). In the formation of the cognitive complexity of an experimental design three variables are involved: the number of scenarios, the number of attributes, and the number of levels. In most cases, these three variables can have different weights during the time required to fill in the questionnaire (Johnson & Meyer, 1984).

Every factorial design, whether fractional or arranged in blocks, draws from a full factorial design. If we define the subset of profiles that form the choice set as S_r, $r = 1, \ldots, q$, the number of factors of each profile as C_i, $i = 1, \ldots, n$, and the number of levels of each factor *as* W_j, $j = 1, \ldots, m$, the number of Johnson and Meyer's cognitive steps would be determined by $S_r \Pi W_j$, Π being the product from j to q levels. Since time is related to effort (Grether & Wilde, 1984), the time required to evaluate the choice set will depend on the number of cognitive steps. Moreover, in the case that all the factors have the same number of levels, the time taken can be expressed as a function of the three elements: time $= f(W_j, C_i, S_r) = W_j \times C_i \times S_r$. Nevertheless, if we apply logarithmic transformation the model becomes additive:

$$T_j = \sum_{p=1}^{P} \beta_p y_{jp} \tag{6}$$

where T_j is the Napierian logarithm of time, β_p is the part-worth of variable p, indicating the slope, and y_{jp} is the Napierian logarithm of the three independent variables.

Given the diversity of subjects and ways to present profiles to respondents (described by text, by paragraphs, by pictures, etc.) and the variety of contexts in which the research can be based (Payne, 1982), it is reasonable to think that a way to estimate the weight each item has in the experiment's duration is by trial and error, because the models estimated in one context can hardly be appropriate in another context (Johnson & Meyer, 1984). In the present study, we follow this logic and we propose response surface methodology (RSM) as an algorithm to optimize the cognitive effort in an experimental design.

RSM is a statistical technique used in the development, improvement, and optimization of industrial processes and it has recently been incorporated into market research studies (Huertas-Garcia, Gázquez-Abad, Martinez-Lopez, & Esteban-Millat, 2013). The methodology was proposed in the early 1950s by Box and Wilson (1951), who justified its use on the basis of the need for efficient experimental procedures that were able to determine the operating conditions of a set of controllable variables. It was considered as a methodology with which an optimal response could be achieved (Box & Draper, 1987). However, it was not developed until the 1970s when some statistical restrictions were overcome and the use of software packages for calculation became widespread (Myers et al., 2009).

RSM is an experimental process that involves sequential stages in which the information obtained in the first stage serves for planning and executing the following stages (Raghavarao, Wiley, & Chitturi, 2011). Nowadays, with the development of market research on the Internet, these sequential models are particularly suitable because they allow for programming questions based on previous answers and, also, can be tailored in real time (Netzer & Srinivasan, 2011; Scholz et al., 2010). The experimental sequence of RSM consists of several stages, beginning with a first-order model

Table 1
Codification of factorial design 2^3.

	Variables	−1	0	+1
1	Profiles	6	8	10
2	Attributes	4	6	8
3	Levels	2	3	4

(i.e. a flat representation) to delimit the slope toward the optimal region and, next, further experiments are developed until reaching the optimum. The algorithm can be summarized in the following five steps (Myers et al., 2009):

1. A first exploratory experimental design of an orthogonal type is proposed and is fitted with a first-order model. For this, two-level vector encoding including a centre point is recommended.
2. Steps are calculated using the first stage's estimated values and 'the fastest route to the minimum' is delimited.
3. Several experiments are carried out throughout the process, noting that the response values decrease, until reaching a point where they begin to grow again; that is, until reaching a turning point.
4. In the case that greater precision were required, the turning point is taken as the basis for developing new experiments to better fit the trend. The adjustment model may continue to be a first-order model.
5. However, if it is noticed that the curvature degree is high, using a second-order model is recommended.

The procedure is designed to achieve independent variable combinations that allow for obtaining highly accurate optimum response values. However, market studies usually do not require such extreme degrees of accuracy given the subjective nature of responses and, therefore, with few experiments it is possible to achieve reasonably accurate approximations to the optimum.

Experiment

The experiment was inspired by the work of Netzer and Srinivasan (2011), who propose the assessment of digital camera profiles (Table 1 shows the attributes and levels considered), although they use a completely different methodology to ours. While they rely on an adaptive self-explicated model for measuring part-worth attributes (a compositional model), in our case we propose the RSM algorithm (a type of decompositional model). In order to arrange the experimental design we departed from a fractional factorial design, specifically from a half factorial design (as can be seen in the first four columns of Table 2). This experimental design has enough variations of profiles, attributes, and levels to contribute to a significant estimation of the factors' part-worth using, as a response variable, the time taken by respondents to assess the camera choice set (Table 3).

In the experiment, we considered three-factor combinations according to the experimental design to gather data from respondents' time spent on making their assessments. In addition, we also manipulated all three factors (profiles, attributes, and levels) around their mean value. These items of information were taken from the work of Netzer and Srinivasan (2011), in which the average number of profiles was 8, for attributes it was 6, and for levels it was 3. This sets a 2^3 factorial experiment around the average point, that is, nine experiments (8 derived from the factorial design plus the central point), and all of them are encoded in vector form (the coding arrangement of the three factors is shown in Table 1). However, due to resource constraints, we have estimated a half factorial design, 2^{3-1} plus the centre point, i.e. 5 profiles. The literature recommends thinking about this first result as the point of origin, 'as a confirmation test, to ensure that conditions experienced during the original experiment have not changed' (Myers et al., 2009, p. 186) (Table 2).

Table 2
Attributes and levels of digital cameras.

Attributes	Levels
1. Brand	Canon, Nikon, Olympus, Sony
2. Battery life	150, 300, 450, 600 photographs
3. Internal memory	8 MB, 16 MB, 32 MB, 64 MB
4. LCD Size	3.81 cm, 5.08 cm, 6.35 cm, 7.62 cm
5. Optical zoom	2×, 3×, 4×, 5×
6. Price	€500, €400, €300, €200
7. Resolution	2, 3, 4, 5 megapixels
8. Warranty	6 months, 1, 2 or 3 years

Source: Adapted from Netzer and Srinivasan (2011).

In order to gather data, a sample of 250 undergraduate students from a large university in Barcelona was used, it was randomly divided into 5 groups and 234 valid questionnaires were obtained (45–48 per group). Each respondent was invited to a computer room and, following researchers' instructions, evaluated a set of digital camera profiles, to which attributes and levels were assigned following the experimental design in Table 2. At the begin of this exercise, respondents had to indicate the exact time registered on the computer clock, then make their assessment of the choice set of digital cameras using a rating score of 1–10 (1 being the least preferred and 10 the most preferred), after their evaluation they had to indicate, once again, the exact time on the questionnaire, and, finally, fill in some identification information and other control questions. Among the control questions, one asked respondents their assessment of the study and another was related to what they thought the objective of the study was. In order to arrange all the different choice set scenarios, which ranged from 6 to 10 profiles, the SPSS orthogonal design generator was used. The study was carried out between February and March 2014. No respondents thought that the study's objective was to measure the time taken; all of them thought it was for measuring their preferences in the digital camera set.

As a dependent or response variable, we used the stated time taken to perform the assessment of the digital camera choice set. To make the calculation easier, the time measured in 60 s was transformed into 100 s, we then calculated the average time of each group, measured in centesimal seconds. Next, this was converted into its natural logarithm and, finally, to calculate the part-worth values the table of contrast coefficients was used. This is a very simple method for estimating the slope of the main factors, given that when the vectorial coding system used with independent variables (a combination of −1 and 1) is transformed into the moment matrix

Table 3
Experiment results.

Experiment number	Profiles	Attributes	Levels	Mean time (in centesimal seconds)	Standard deviation
2	1	−1	−1	150.378629	50.6857802
3	−1	1	−1	138.75024	50.5166211
5	−1	−1	1	58.4159032	32.3333863
8	1	1	1	362.383978	97.5260192
9	0	0	0	157.857452	49.5892357

Table 4
Contrast coefficients calculation table.

Experiment number	Constant	Profiles	Attributes	Levels	Nap-log mean time
2	5.01315631	5.01315631	−5.01315631	−5.01315631	5.01315631
3	4.93288265	−4.93288265	4.93288265	−4.93288265	4.93288265
5	4.06758817	−4.06758817	−4.06758817	4.06758817	4.06758817
8	5.89270436	5.89270436	5.89270436	5.89270436	5.89270436
9	0	0	0	0	5.06169242
Sum	19.9063315	1.90538985	1.74484254	0.01425357	
Divisor	4	4	4	4	
Average	4.97658287	0.47634746	0.43621063	0.00356339	
Standard error	0.03442457	0.06884913	0.06884913	0.06884913	

it becomes an identity matrix and, therefore, the regression model to fit is simply $y = \mathbf{Xb}$. In other words, the coding matrix is multiplied by the vector of results to obtain a finding matrix, then, by adding the values of each column and dividing them by the number of positive and negative coded levels, it is possible to obtain the part-worth values as though adjusted by OLS (Table 4 illustrates the calculations). In order to calculate the standard error, we followed the process of estimating the average variance proposed by Box et al. (2005).

The first-order model derived from the experiment is as follows:

$$\hat{T}_j = 4.976 + 0.476y_1 + 0.436y_2 + 0.003y_3 \tag{7}$$

where T_j is the Napierian logarithm of average time spent on the assessment process, y_1 is the number of profiles, y_2 is the number of attributes, and y_3 is the number of levels. As we can see in Table 4, only the number of profiles and the number of attributes have a significant impact on the time used for assessment and both have a similar weight ($\beta_1 = 0.476$ and $\beta_2 = 0.436$), while the number of levels does not seem to have any effect on the amount of time used by respondents.

Once the model has been fitted and the weight of the main factors has been estimated (7), we need to set the path to achieve the minimum value of T_j. To calculate the path to reach the minimum we followed the algorithm proposed by Myers et al. (2009):

In the first stage, we need to estimate the length of the steps that will be taken to reach the minimum; usually the length of the step is 1, say Δx_i. This length is assigned to the independent variable whose part-worth is the highest in absolute value, in our case it is ($\beta_1 = 0.476$).

We then estimate the length of the other variables' steps, whose relative longitude will depend on the weight of each factor, following this function:

$$\Delta x_j = \frac{\beta_j}{\beta_i / \Delta x_i}, \quad j = 1, 2, \ldots, k, i \neq j \tag{8}$$

$$\Delta x_2 = \frac{0.436}{0.476} = 0.915$$

$$\Delta x_3 = \frac{0.003}{0.476} = 0.007$$

Finally, coded variables are transformed into their normal values. Each part-worth measures the slope degree that represents a change from 0 to 1 in the rating scale. For example, β_1 represents the change from 0 to 1 that meant a change from 8 to 10 profiles, i.e. an increase of 2 profiles. In the case of β_2, the slope represents the change from 6 to 8 attributes, i.e. an increase of 2 attributes, and, finally, β_3 represents the change from 3 to 4 levels, this time with an increase of 1 level. As defined above, the step length for the main important factor (number of profiles) is 1, and this step generated an increase of 0.476 in the amount of time, which corresponds to a change in two profiles. For the other independent variables, the length of their steps is proportional to the values resulting from Eq. (3), as shown in Table 5.

In many cases, it is possible to find a good approximation to the optimum simply with the first simulation. Particularly in the case when all values are discrete levels, as in this study. As shown in Table 5, with the 'base – 2 steps', the simulation table finds a good approximation to the optimal solution. This is obtained with the combination of 4 profiles with 2 attributes each and 3 levels per attribute, and to complete this assessment respondents take 25.15 centesimal seconds ($e^{3.22} = 25.15$). In the following simulation, the 'base – 3 steps', the model suggests an impossible solution because in the combination of profiles, attributes, and levels it considers less than one attribute in the choice set. However, if instead of working with discrete values we were working with continuous values, it could be possible to obtain greater accuracy, simply by developing new experiments around the last point estimated. It could also be possible to use a second-order equation to fit the model, trying to cover the degree of curvature that is usually near the optimum.

In short, developing an experimental design to carry out CA research is not an easy task as it must fulfil two contradictory objectives: firstly, achieve statistical efficiency, helping to improve the parameters' estimation, and, secondly, ensure effective and trusted answers from respondents taking into account their limited cognitive capability. With these arguments, using software for generating experimental designs is not a panacea, because the type of protocol used to collect data and the context in which the experiment is developed determine the behavioural pattern followed by consumers in evaluating profiles (Johnson & Meyer, 1984; Payne, 1982; Payne et al., 1988). Therefore, according to the advice of Kuhfeld et al. (1994), the best strategy for building an

Table 5
Simulation process of RSM algorithm.

	Coding variables			Natural variables			Nap-log time*
	Profiles	Attributes	Levels	Profiles	Attributes	Levels	
Base	0	0	0	8	6	3	5.06169242
Step length	1	0.91574044	0.00748066	2	1.83148088	0.00748066	
Base – 1 step	−1	−0.91574044	−0.00748066	6	4.16851912	2.99251934	4.10075303
Base – 2 steps	−2	−1.83148088	−0.01496132	4	2.33703824	2.98503868	3.2249232
Base – 3 steps	−3	−2.74722132	−0.02244198	2	0.50555737	2.97755802	2.34909336

* Values estimated from Eq. (2).

experimental design is to combine computer tools, which provide designs quickly and easily, with the researcher's insight to select the most appropriate design option for the context of the study.

Conclusions

CA has been one of the most common market research techniques used by researchers and practitioners to measure consumer preferences for more than four decades. This technique emerged to overcome some shortcomings in the use of self-explicated questionnaires, which had been criticized for being vague and unrealistic (Hauser & Rao, 2004; Sattler & Hensel-Börner, 2007). To implement CA, a fundamental step is building the experimental design, which determines the number of profiles that form the choice set and the pattern for arranging them by combining attributes and levels. To make these designs statistically efficient, they must be orthogonal and balanced, and this is only possible with very simple designs; i.e. those which require few variables and few levels. Therefore, most experimental designs are quasi-orthogonal and, in this case, it is necessary to measure their degree of statistical efficiency.

There are several ways to measure statistical efficiency. In this paper, we have reviewed the two most common optimization criteria in the literature: A-Optimal and D-optimal criteria. However, these criteria are based on the underlying assumption that an algebraic relationship between product attributes and utility explain consumer behaviour, and that this relationship is independent of the number of alternatives and of the experimental context. These assumptions have been criticized as unrealistic from the perspective of consumer cognitive behaviour theories. To overcome these cognitive limitations a cognitive efficiency criterion is proposed which can act as a complement to the statistical efficiency criterion. In addition, we have also proposed a function to estimate this efficiency and an algorithm to optimize it. The function is inspired by the cognitive steps proposed by Johnson and Meyer (1984), where the number of profiles, factors, and levels are independent variables, and the length of time to fill in the questionnaire is the dependent variable (Grether & Wilde, 1984). Using the time taken for the choice set assessment as an estimate of cognitive effort is completely appropriate since market research uses the Internet and websites where respondents' patience tends to be low (Deutskens, de Ruyter, Wetzels, & Oosterveld, 2004) and their willingness to participate decreases as the length of the survey extends (Scholz et al., 2010). In addition, the algorithm proposed to optimize this function is RSM and to illustrate this whole process we carried out an experiment using digital cameras as stimuli for respondents to choose among them.

This study responds to the request made by Bradlow (2005) to develop techniques to assess consumer preferences, allowing, at the same time, for the evaluation of a large number of attributes without cognitively burdening respondents. Its contribution to the literature is threefold: firstly, experimental designs are conceptualized using both statistical and cognitive efficiency; secondly, a functional relationship between all the components of an experimental design (profiles, factors, and levels) is established, as well as an estimator of cognitive effort as target function; and thirdly, a sequential and dynamic algorithm is proposed that allows the objective function reach an acceptable optimum.

Hybrid proposals have emerged, such as adaptive conjoint analysis, with the aim of balancing statistical efficiency and cognitive effort. However, the fact of using partial profiles during the compensatory phase has been criticized by both scholars and market researchers for the inability to obtain stable factor estimations in individual studies (Orme, 2007). There have been some proposals (for example, Huertas-Garcia, Gázquez-Abad, & Forgas-Coll, 2016) in which partial profile scenarios are replaced with full profiles, but this solution is only viable for a small number of factors.

One limitation of this study is that the analysis has only focused on classical CA without considering discrete choice experiments, which are characterized by each respondent only choosing the best option from the choice set. Although choice has been defined as much more natural behaviour in consumption than ranking and rating (Louviere, Hensher, & Swait, 2007), the items of information gathered by the researcher from each respondent are minimal (1, if x alternative was chosen from the entire choice set, and 0 otherwise), therefore its use is restricted to studies that use very large samples. Another limitation in our study is that only the existence of a compensatory model in consumer choice processes was considered, regardless of the existence of other non-compensatory heuristic models (lexical, qualifying, etc.) that could explain the process of choosing among many alternatives.

Furthermore, due to lack of resources, the experiment was carried out with the following restrictions: firstly, it was performed once, so more replicas are needed to verify the results; secondly, in the initial phase of RSM only a half factorial design was developed and with this restriction it was not possible to estimate all two-factor interactions, which would have contributed to more interesting results; thirdly, the sequential experiments needed to follow the steps to reach the minimum were simulated assuming a linear trend between the factors. Therefore, possible further research could involve conducting new tests using a full factorial design in the first phase and carrying out all the necessary steps until reaching the optimum.

References

Beusterien, K., Grinspan, J., Kuchuk, I., Mazzarello, S., Dent, S., Gertler, S., et al. (2014). Use of conjoint analysis to assess breast cancer patient preferences for chemotherapy side effects. *The Oncologist*, *19*(2), 127–134.

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York, USA: John Wiley & Sons.

Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters: Design, innovation, and discovery* (2nd ed.). Hoboken, USA: John Wiley & Sons.

Box, G. E., & Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society, Series B (Methodological)*, *13*(1), 1–45.

Bradlow, E. T. (2005). Current issues and a 'wish list' for conjoint analysis. *Applied stochastic models in business and industry*, *21*(4–5), 319–323.

Chrzan, K., & Orme, B. (2000). An overview and comparison of design strategies for choice-based conjoint analysis. In *Sawtooth software conference proceedings*. Sequim, WA: Sawtooth Software Inc.

Dellaert, B. G., Donkers, B., & Soest, A. V. (2012). Complexity effects in choice experiment-based models. *Journal of Marketing Research*, *49*(3), 424–434.

DeShazo, J. R., & Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, *44*(1), 123–143.

Deutskens, E., De Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters*, *15*(1), 21–36.

East, R. (1973). The duration of attention of alternatives and re-evaluation in choices with two and three alternatives. *European Journal of Social Psychology*, *3*(2), 125–144.

East, R., Wright, M., & Vanhuele, M. (2013). *Consumer behaviour: Applications in marketing*. London, UK: Sage.

Green, P. E., Krieger, A. M., & Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, *31*(3 Suppl.), S56–S73.

Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, *54*(4), 3–19.

Grether, D., & Wilde, L. (1984). An analysis of conjunctive choice: Theory and experiments. *Journal of Consumer Research*, *10*(4), 373–385.

Gustafsson, A., Herrmann, A., & Huber, F. (2007). Conjoint analysis as an instrument of research practice. In A. Gustafsson, A. Herrmann, & F. Huber (Eds.), *Conjoint measurement. Methods and applications* (pp. 3–30). Berlin, Germany: Springer.

Hauser, J. R. (2014). Consideration-set heuristics. *Journal of Business Research*, *67*(8), 1688–1699.

Hauser, J. R., & Rao, V. R. (2004). Conjoint analysis, related modeling, and applications. In Y. Wind, & P. Green (Eds.), *Marketing research and modelling: Progress and prospects* (pp. 141–168). New York, USA: Springer-Verlag US.

Huertas-Garcia, R., Gázquez-Abad, J. C., & Forgas-Coll, S. (2016). A design strategy for improving adaptive conjoint analysis. *Journal of Business & Industrial Marketing*, *31*(3), 328–338.

Huertas-Garcia, R., Gázquez-Abad, J. C., Martínez-López, F. J., & Esteban-Millat, I. (2013). Using response surface methodology to optimize factors in conjoint experiments. *International Journal of Market Research*, *55*(2), 267–288.

Johnson, R. (1987). *Accuracy of utility estimation in ACA* Working Paper. Sequim, WA: Sawtooth Software.

Johnson, E. J., & Meyer, R. J. (1984). Compensatory choice models of noncompensatory processes: The effect of varying context. *Journal of Consumer Research*, *11*(1), 528–541.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society Series B*, *21*, 272–304.

Kuhfeld. (1997). Efficient experimental designs using computerized searches. In *Proceedings of the Sawtooth Software conference* (pp. 71–86). Sequim, WA: Sawtooth Software Inc.

Kuhfeld, W. F., Tobias, R. D., & Garratt, M. (1994). Efficient experimental design with marketing research applications. *Journal of Marketing Research*, *31*(4), 545–557.

Lancaster, K. (1990). The economics of product variety: A survey. *Marketing Science*, *9*(3), 189–206.

Louviere, J. J. (2001). What if consumer experiments impact variances as well as means? Response variability as a behavioral phenomenon. *Journal of Consumer Research*, *28*(3), 506–511.

Louviere, J. J., Hensher, D., & Swait, J. (2007). Conjoint preference elicitation methods in the broader context of random utility theory preference elicitation methods. In A. Gustafsson, A. Herrmann, & F. Huber (Eds.), *Conjoint measurement: Methods and applications* (4th ed., pp. 167–198). Berlin, Germany: Springer.

Louviere, J. J., Islam, T., Wasi, N., Street, D., & Burgess, L. (2008). Designing discrete choice experiments: Do optimal designs come at a price? *Journal of Consumer Research*, *35*(2), 360–375.

Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2009). *Response surface methodology* (3rd ed.). Hoboken, USA: John Wiley & Sons Inc.

Netzer, O., & Srinivasan, V. (2011). Adaptive self-explication of multiattribute preferences. *Journal of Marketing Research*, *48*(1), 140–156.

Orme, B. K. (2007). Which conjoint method should I use? In *Sawtooth Software research paper series*. Sequim, WA: Sawtooth Software Inc.

Olshavsky, R. W. (1979). Task complexity and contingent processing in decision making: A replication and extension. *Organizational Behavior and Human Performance*, *24*(3), 300–316.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, *92*(2), 382–402.

Payne, J. W., Bettman, J. R., & Johnson, E. L. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552.

Raghavarao, D., Wiley, J. B., & Chitturi, P. (2011). *Choice-based conjoint analysis. Models and designs* (1st ed.). Boca Raton, USA: Taylor and Frances Group.

Salisbury, L. C., & Feinberg, F. M. (2010). Alleviating the constant stochastic variance assumption in decision research: Theory, measurement, and experimental test. *Marketing Science*, *29*(1), 1–17.

Sattler, H., & Hensel-Börner, S. (2007). A comparison of conjoint measurement with self-explicated approaches. In A. Gustafsson, A. Herrmann, & F. Huber (Eds.), *Conjoint measurement: Methods and applications* (4th ed., pp. 67–76). Berlin, Germany: Springer.

Scholz, S. W., Meissner, M., & Decker, R. (2010). Measuring consumer preferences for complex products: A compositional approach based on paired comparisons. *Journal of Marketing Research*, *47*(4), 685–698.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*, 1–19.

Toubia, O., & Hauser, J. R. (2007). On managerially efficient experimental designs. *Marketing Science*, *26*(6), 851–858.

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

Vermeulen, B., Goos, P., & Vandebroek, M. (2008). Models and optimal designs for conjoint choice experiments including a no-choice option. *International Journal of Research in Marketing*, *25*(2), 94–103.