The Australian Journal of Agricultural and Resource Economics

Journal of the Australian Agricultural and Resource Economics Society

The Australian Journal of Agricultural and Resource Economics, 52, pp. 253–282

Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why*

Riccardo Scarpa and John M. Rose†

We review the basic principles for the evaluation of design efficiency in discrete choice modelling with a focus on efficiency of WTP estimates from the multinomial logit model. The discussion is developed under the realistic assumption that researchers can plausibly define a prior belief on the range of values for the utility coefficients. *D*-, *A*-, *B*-, *S*- and *C*-errors are compared as measures of design performance in applied studies and their rationale is discussed. An empirical example based on the generation and comparison of fifteen separate designs from a common set of assumptions illustrates the relevant considerations to the context of non-market valuation, with particular emphasis placed on *C*-efficiency. Conclusions are drawn for the practice of reporting in non-market valuation and for future work on design research.

Key words: efficient experimental design, WTP-efficiency, C-efficiency, experimental design, choice modelling, choice experiments.

1. Introduction

Stated choice modelling has now an established role in non-market valuation. Practitioners are engaged in testing the method and defining the boundaries of its use in public decision making and cost benefit analysis. In this respect the method has taken up a research agenda which is quite distinctive from other fields of applications, such as in transport, marketing, food choice and health research. One of the areas of distinctiveness is associated with the methodology of experimental design for the specific purpose of deriving non-market values

A survey of existing non-market valuation studies indicates that there is a prevailing format of stated choice surveys in non-market valuation. Typically, these surveys involve asking respondents to indicate their preferred alternative from those offered within a given choice set. Alternatives in the choice set are often outcomes of policies that can vary in their effects of relevance to the respondent. Effects of policies are described by a selected number of attributes, each of which can take a qualitative or numerical level. Rather

* We would like to thank the anonymous reviewers of this journal for useful comments and the financial support of the Waikato Management School competitive funds award.

† Riccardo Scarpa (email: rscarpa@mngt.waikato.ac.nz), University of Waikato, Hamilton, New Zealand and John M. Rose (email: johnr@itls.usyd.edu.au), University of Sydney, Sydney Australia.

doi: 10.1111/j.1467-8489.2007.00436.x

than reviewing a single choice set, respondents are typically asked to evaluate several choice sets, thus increasing the number of observations per individual surveyed and producing a panel of discrete choice responses. Underlying stated choice surveys are experimental designs, which are used to allocate the levels of the attributes that make up the alternatives within each of the choice sets used in the survey. As such, experimental designs lie at the core of all stated choice studies.

Conceptually, experimental designs may be viewed as the systematic arrangement in matrices of the values that researchers use to describe the attributes representing the alternative policy options of the *hypothetical* choice sets. As the total number of possible combinations of attribute and attribute levels can be huge even with relatively simple problems, some theory must be used to drive the selection of these levels and their arrangements in the choice sets to achieve the required information within practical sample sizes.

Via experimental design theory, the analyst is able to determine the values to be assigned to attributes in each alternative situated within the choice sets to be used in the survey. The assignment of these values occurs in some systematic (i.e. non-random) manner so as to achieve the intended results of the survey in an efficient (i.e. a least cost) manner. Cost effectiveness is of paramount importance in many non-market valuation studies, so design efficiency is a much sought after property as it allows researchers to minimise the sample sizes necessary to achieve a given degree of estimation accuracy. Design theory makes use of various criteria to evaluate the outcomes of these assignments on the basis of the assumptions invoked by the analyst as incorporated by a given model specification. The selection of the correct set of criteria will drive the analyst to an adequate choice of experimental design for the purpose at hand. However, this will be conditional on the chosen specification and on other necessary assumptions made by the researcher.

Experimental design techniques are of general relevance in survey research. However, the specific focus of non-market valuation on the derivation of implicit prices from discrete choices has some important and distinctive implications in experimental design practice. Such implications are still inadequately addressed in the literature, as recently noted, for example, by Ferrini and Scarpa (2007) and Toubia and Hauser (2007). The present paper intends to contribute to developing an understanding of these implications within the 'workhorse' of discrete choice analysis: the conditional logit model predicated on random utility theory. Extensions to other specifications of the logit family are conceptually immediate, although technically challenging, and definitely beyond the scope of this paper.

To do so, we selectively draw from the wide and rapidly expanding literature in experimental design for logit models and we propose an unfrequently used criterion based on the specific needs of non-market valuation. For choice modelling surveys developed to estimate monetary values, desirable criteria should revolve around efficiency of willingness to pay (WTP) estimates. For models specified in the preference space, which represent the vast majority

used in practice,[1] WTP for single attributes are functions of parameter estimates of logit models predicated on random utility theory. While criteria measuring predictive performance of probabilities, utility balance across alternatives and efficiency of the utility estimates are much more frequently used in design evaluation, the way such criteria are related to efficiency of and sample size requirements for WTP estimates is unclear. In this paper we set up the building blocks for investigating such a relationship and provide a worked out example exploring the relationship between parameter efficiency, WTP efficiency and sample size requirements for stated choice surveys. We set up our example in a setting that is most common in non-market valuation applications, the one with repeated choices from two hypothetical alternatives and the status-quo or no-buy option. These are the most frequent operational conditions in non-market valuation studies.

The rest of the paper is organised as follows. Section 2 provides a discussion of the relationship between discrete choice models, random utility theory and experimental design. Section 3 discusses various efficiency criteria that have been employed in the literature before the introduction of an uncommon criteria based on WTP efficiency, which is discussed in Section 4. Section 5 provides a brief discussion on what should be reported in terms of statistical measures after which Section 6 discusses various algorithms for generating efficient designs. In Section 7 we discuss the issue of scaling, which has implications for designs with status-quo and it has often been ignored by the existing literature. In Section 8 we discuss a case study in which 15 experimental designs optimised on the basis of various criteria and generated using different design strategies are contrasted. Our conclusions and ideas for further research are reported in Section 9.

2. Discrete choice models, random utility and experimental design

Qualitative choices are based on discrete outcomes represented by the selection of alternatives from given consideration sets. What form of evaluation (lexicographic, elimination by aspect, economic or other attribute screening rules, etc.) is predominant among respondents in driving such selections remains an elusive issue. Much research is being conducted on methods to practically distinguish these processes starting from observed behaviour. Regardless of actual evaluation processes, in applied research, the most successful paradigm to date has been the random utility theory (RUT), and we refer to this in what follows. Similarly, in terms of statistical analysis of responses, the most successful specification consistent with RUT has been the conditional logit model (McFadden 1974). This model remains at the core

[1] Random utility applications can accommodate models directly specified in the WTP space. For examples of this kind from stated preference data see Train and Weeks (2005), and for an example from revealed preference data see Scarpa *et al*. (in press). In these specifications attribute WTPs are coefficient estimates and not derived from functions of such estimates.

of most of the more sophisticated specifications, such as nested and mixed logit models. What is discussed and illustrated in practice here can be easily extended, although not so easily illustrated, to more sophisticated RUT-based models.

The main point of departure of our study concerns the logical consequences from being able to assume the direction and sometime the relative magnitude of the values of the taste intensity parameters in the utility function. Unfortunately, within the published literature there appears to have developed two separate paradigms for constructing designs, both of which claim to generate designs aimed at improving the overall efficiency of stated preference surveys. Regrettably, both paradigms use different criteria to judge the overall level of efficiency of a generated experimental design. One design paradigm seeks to maximise differences between the attribute levels of the stated preference alternatives, whereas the second method attempts to minimise the variances of the parameter estimates obtained for each of the attribute coefficients included in the utility specification. This second approach is based on the observation that as soon as the researcher can plausibly defend that some attributes of choice may plausibly be expected to have a given sign or relative size, the efficiency of the design for a logit specification can easily be shown to be improved from what would be the case in the absence of such assumptions. Because of these radically different starting assumptions our work is to be located in the second paradigm and cannot be compared to similar research carried out within the limited framework of probability balanced designs, that are predicated on researchers' total ignorance of the values of taste intensities (e.g. Street *et al.* 2001, 2005; Burgess and Street 2005; Lusk and Norwood 2005; Street and Burgess 2005). With this premise, these authors proceed to develop a discussion prevalently based on the property of orthogonality,[2] which is – as they themselves note – much more relevant for designs developed for linear multivariate models than it is for highly non-linear models such as those in the logit family.

As a matter of fact, we and many others (e.g. Sándor and Wedel 2001, 2002, 2005; Bliemer and Rose 2005; Bliemer *et al.* in press; Ferrini and Scarpa 2007 and Kanninen 2002; Kessels *et al.* 2006) argue exactly the opposite, which is that in the greatest majority of non-market valuation studies, researchers indeed *are* able to predict at least the sign of the price coefficient. In reality, however, researchers can normally do more than this and express some beliefs on the range of (relative) values that are likely to be taken by other parameters in the utility function.

In terms of assumptions our research is therefore more akin to research efforts by Sándor and Wedel (2001, 2002, 2005), Bliemer and Rose (2005), Bliemer *et al.* (in press), Ferrini and Scarpa (2007) and Kessels *et al.* (2006). We also note that this approach is more in keeping with previous literature in optimal design for non-market valuation (Kanninen 1993a,b; Alberini 1995), and of sequential improvement of survey designs in non-market valuation

[2] Orthogonality here refers to experimental designs (or data) where the attributes of the design (or data) are uncorrelated with one another.

(Kanninen 1993b; Scarpa *et al*. 2007). It is also more consistent with the notion that selected attributes for choice modelling studies are usually relevant to respondents and hence different from zero in the population of interest.

We will show with examples that when adequately expressed, this *a priori* information is of great use and can lead to substantial efficiency gains in the design. In doing so, however, the analyst must be made aware of some potential difficulties, some of which are of specific interest to the current choice modelling practice for the purpose of non-market valuation, such as the effect of the status-quo alternative and that of the choice of attribute coding on the evaluation of the efficiency of the design.

We now move our attention to the definition of efficiency in the context of the logit model commonly used to derive estimates of utility coefficients from observed discrete choice.

3. Measuring design efficiency for taste intensities

In this section we examine the measures of design efficiency that are of interest when the objective is to estimate the coefficients of the indirect utility function, or the so-called 'taste intensities'.

3.1 The basics

Consider a situation involving the choice between $i, j = 1, 2, \ldots, J$ alternatives, each of which is described by $k = 1, 2, \ldots, K$ attributes. Assuming the choice process of choice situation $s = 1, 2, \ldots, S$ is modelled using a conditional logit specification with Gumbel error scale $\lambda > 0$, we get:

$$\Pr(Y_s = i) = \frac{e^{\lambda \beta' x_{si}}}{\sum_{j=1}^{J} e^{\lambda \beta' x_{sj}}}, \quad \lambda > 0, \tag{1}$$

as the probability that alternative i will be selected from the set of J alternatives available in choice task s.

The specific values of x_{sj} are defined by the experimental design. An efficient design will minimise the variance-covariance estimator, or – put differently – will maximise the amount of information the design conveys to identify the estimates of the vector β. The information matrix for the design under the conditional logit assumption is given by the matrix of second derivatives of the log-likelihood function, which can compactly be written as:

$$I(\beta, x_{sj}) = \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\sum_{n=1}^{N}\sum_{s=1}^{S}\sum_{j=1}^{J} P_{sj}(x_{sj} - \bar{x}_{sj})(x_{sj} - \bar{x}_{sj})' \tag{2}$$

$$\text{with} \quad \bar{x}_{sj} \equiv \sum_{j=1}^{J} P_{sj} x_{sj},$$

where n denotes respondent $n = 1, 2, \ldots, N$ and s choice situation $s = 1, 2, \ldots, S$. The resulting matrix will be of size $K \times K$.

One of the reasons for the popularity of the multinomial logit model is that of having a relatively simple mathematical formulation of both the Jacobian (gradient or vector of first derivatives) and Hessian (matrix of second derivatives) models. These two objects, however, are functions of both the utility coefficients β and the matrix of choice attributes x_{nsj} (i.e. the experimental design). So, an informative design is one that increases the size of the elements contained within $I(\beta, x_{nsj})$. In other words, taken $g(I(\beta, x_{nsj}))$ as a measure of information, an informative design should make this measure large. At this stage it is useful to revise the relationship between $I(\beta, x_{nsj})$ and a common Maximum Likelihood (ML) estimator of the asymptotic variance-covariance (AVC) matrix $\Omega(\beta, x_{nsj})$ of a design. The Maximum Likelihood estimator of the AVC matrix for a design to be used with the conditional logit model is the negative of the inverse of the expected Fisher information matrix (e.g. see Train 2003), where the latter is equal to the second derivatives of the log-likelihood function:

$$\text{AVC} = \Omega(\beta, x_{nsj}) = [E[I(\beta, x_{nsj})]]^{-1} = \left[-\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \right]^{-1}, \tag{3}$$

where $\ln L$ is the log-likelihood of the design:

$$\ln L = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} y_{nsj} \ln P_{nsj}\left(x_{nsj}, \beta\right), \tag{4}$$

and s denote the choice tasks implied in the design, n the respondent, y_{nsj} the choice indicator, while j denote the alternatives.

Because of this inverse relationship, in choosing an informative (efficient) design, one can choose to think in equivalent terms of either *maximising* information or *minimising* variance. A suitable algorithm would search the arrangement of attribute and levels, in a suitably coded matrix x_{nsj}, such that an optimal solution is found according to some stopping criteria from the sometime extremely large feasible set of solutions.

3.2 Design efficiency measures

A key passage is the definition of the function $g(\cdot)$, which is useful to define as a single number, rather than a collection of numbers, as in vectors and matrices. A convenient scalar measure of the size of a matrix is its determinant, which is a sum of terms, each made-up of products of systematically selected elements of the matrix. A non-zero determinant matrix implies that the matrix has full rank (no collinearity and identification of the β). So, the determinant of the information matrix (or equivalently minimising that of the AVC) is a valid measure of the efficiency of a candidate design. However,

the determinant will be larger as K – the number of elements in β – increases, so that one must devise a measure that accounts for that too. An often used measure is the *D*-error:

$$D\text{-error} = \det(\Omega(\beta, x_{sj}))^{1/K} \tag{5}$$

Note that when each respondent is asked to review the same set of choice tasks, it is common to compute such efficiency measures assuming a single respondent. Although not necessary, we currently adopt this assumption for simplicity, and as such, we drop the subscript *n*.

Rather than the determinant, another measure of efficiency has been used the so-called *A*-efficiency, measured by the *A*-error, which is the trace of the AVC:

$$A\text{-error} = \text{trace}(\Omega(\beta, x_{sj})). \tag{6}$$

However, this measure seems to have encountered lower acceptance and use within the published literature. The reason for this is that only the main diagonal elements of the AVC matrix are used in computing the trace, and hence this measure does not account for the off-diagonals and so it may produce very large covariances for the parameters.

One final measure, which we explore in this paper, does not look at the AVC matrix, but at the choice probabilities for the design. This measure, proposed by Kessels *et al*. (2004), is not explicitly meant to be used as a measure of design efficiency, however, we use it here as a means of attempting to prevent choice sets containing alternatives that may be strongly dominated. The probability or utility balance of a design is given by the following statistic:

$$B = \frac{\sum_{s=1}^{S}\left(\prod_{j=1}^{J}\Pr\left(Y_s = j\right)\right)}{S\left(\frac{1}{J}\right)^{J}} \tag{7}$$

Equation (7) will range between zero and 100 per cent, with the percentage value representing how balanced the probabilities (or utilities) are over the alternatives within the design. A zero value indicates that there exists a completely dominant alternative within each choice set, whereas a value of 100 per cent indicates that each alternative in every choice set has an equal probability of being chosen.

In constructing a design, different combinations of attribute levels (i.e. different experimental designs) will produce different *D*- and *A*-error measures, and different B statistic values. Given that lower *D*- or *A*-error values typically correspond to lower values within the AVC matrix under consideration (e.g. the smaller the determinant of a matrix, the smaller on average will be the elements contained in that matrix), locating a design with a lower *D*-error leads to a design that is expected to produce smaller parameter variances and

covariances. On the other hand, the use of the *A*-error as a design criterion would be expected to result in a design with minimum variances, but not necessarily small covariances, this is because the trace only considers the diagonals of the AVC matrix. Thus, searching over different arrangements of attribute levels in x_{sj} with the objective of minimising these scalar measures allows the analyst to search for designs that will be expected to produce smaller standard errors (and covariances depending on the criteria employed). As such, designs with lower *D*- or *A*-errors are typically said to be designs that are *D*- or *A*-efficient. The same is true of other efficiency measures discussed below.

Note that the probability balance of the design, as given by the *B*-statistic in Equation (7), is typically not used in generating efficient designs. This is despite there being a clear relationship between this value and the elements in the AVC matrix of discrete choice models (see Kanninen 2002).[3]

We would argue that any other measure is not only relatively uninformative, but in some cases can even be misguiding. Consider the frequent practice of reporting the following design statistic in stated choice studies:

$$D\text{-efficiency} = \frac{100}{\left[S\left|\left(X'X\right)^{-1}\right|^{1/K}\right]}, \tag{8}$$

(e.g. Lusk and Norwood 2005, p. 772) where *S* is the number of observations (i.e. choice sets), *K* is the number of attributes in the design and *X* the design matrix.

This measure is uninformative with respect to the operating conditions of discrete choice modelling under random utility theory. This is because Equation (8) is derived under the assumption that the model to be estimated is linear in nature. The relationship between this equation and that of the variance-covariance matrix of the homoskedastic linear regression model, $\sigma^2[X'X]^{-1}$ clearly demonstrates the relationship between the two. Unfortunately, the variance-covariance matrices of discrete choice models are radically different to those of linear models as can be seen by inspecting the Hession given in Equation (2), which depends on the values of β. Indeed, Equation (8) will return a value of 100 per cent for an orthogonal[4] design and lower values

[3] Interestingly, the relationship is such that probability balance is unlikely to yield the most efficient design, contrary to the work presented by Huber and Zwerina (1996). Kanninen (2002) demonstrates that in the case of designs with two alternatives, the optimal choice probabilities are related to the number of attributes of the design, with designs producing choice probabilities around 0.7 and 0.3 resulting the smallest possible values for the elements contained within the AVC matrices of discrete choice models.

[4] Note that Equation (8) only works under very strict coding structures. The statistic is only valid if the design is coded using orthogonal coding (i.e. −1, 1). If the design is coded in any other manner (or there are more than two levels – thus requiring more levels than −1 and 1), then the statistic fails. We also note that the formula proposed in Lusk and Norwood (2005, page 772) is more frequently used in its equivalent form: $D\text{-efficiency} = 100 \times |X'X|^{1/K}/N$, and that *K* is not the product of attributes and levels in the design, as suggested by these authors, but the number of attributes.

for non-orthogonal designs. As we argue later however, design orthogonality of this type does not imply efficiency of non-linear discrete choice models.[5]

We note that although we deliberately restricted the discussion to the conditional logit model, these principles are fully applicable to any model of discrete choice, such as the nested logit or mixed logit models. All it requires is the computation of the adequate information matrix (see e.g. Bliemer and Rose 2006).

3.3 Design specificity and coefficient uncertainty

Two important observations are in order here, both of which clearly affect the measurement of efficiency of a conditional logit design. The first concerns the coding of the variables in the matrix x_{sj} and it concerns the fact that efficiency depends on the type of coding chosen (on the levels, effect-coding, or dummy variable coding). As a consequence, a design obtained under effect-coding will produce different efficiency values if the coding is changed to dummy variable coding or to the levels. Hence the efficiency measures should not be compared across models with different coding applied to the same design.

The second issue concerns the assumptions about the values of β, which are the very quest of a stated choice survey study and hence cannot be known with certainty at the time of designing the experiment. These can, however, be assumed by the analyst to be in a given range with some degree of uncertainty. Such uncertainty can be formally defined in terms of adequate *a priori* distributions, as done for example in Sándor and Wedel (2001) and in Ferrini and Scarpa (2007).

For this reason the literature distinguish between *point D*-error and *Bayesian D*-errors, using the notation D_p and D_b, respectively. The latter is just an expectation taken over the assumed *a priori* distributions of β. Suppose, for example that the values of β are *a priori* believed to be distributed normally, with a vector of means μ and a variance-covariance Σ. Then the D_b error would be:

$$D_b \text{ error} = \int \left[\det\left(\Omega\left(\beta, x_{sj} \right) \right) \right]^{1/K} N(\mu, \Sigma) d\beta \tag{9}$$

Of course, less informative priors can be invoked, such as uniform distributions over a broad range of values.

3.4 Level of significance and design replicates

The survey will typically employ many replications of the same design and generally a design will be completed by more than one respondent in the final

[5] To pre-empt our argument, discrete choice models are estimated on the differences of the utilities of the chosen and non-chosen alternatives, and not on the actual data itself. Thus, the orthogonality of the levels of the data matrix is not as important as is orthogonality in the differences of attribute values between alternatives.

data. In generating the design, it is common to assume only a single respondent, however, this need not always be the case. In particular, it is useful to assume more than one respondent when the design is too large for a single respondent and subsets of the design are to be given to different respondents (this is commonly achieved via *blocking* the design). Suppose that a design is broken into G subsets, with n respondents reviewing each subset (noting that n may be different for each G). Then the AVC matrix for the final model would be:

$$\Omega_N\left(\beta, x_{sj}\right) = \sum_{g_n=1}^{G_n} \Omega_{g_n}\left(\beta, x_{sj}\right) \tag{10}$$

Further, note that the AVC has dimension $K \times K$ and that the asymptotic standard errors for each estimate of the elements of β are given by the squared root of the diagonal of the AVC matrix:

$$\begin{bmatrix} s.e._1 \\ s.e._2 \\ \vdots \\ s.e._k \end{bmatrix} = \sqrt{\operatorname{diag}\left(\Omega\left(\beta, x_{sj}\right)\right)} \tag{11}$$

This is sometimes used to derive a measure of the (theoretically) required design replicates to achieve a given significance value for a choice attribute coefficient k via the required t-value and the relationship[6]:

$$t_{\beta_k} = \frac{\beta_k}{s.e._k}\sqrt{n_{\beta_k}} \rightarrow n_{\beta_k} = \left[\frac{t_{\beta_k} s.e._k}{\beta_k}\right]^2 \tag{12}$$

For example, suppose one assumes a $\beta_1 = 1.2$ and derives a design with an $s_1 = 2$ but wants to compute the number of design replicates necessary to achieve a five per cent significance for which the two-tailed t-value is approximately 1.96. Then an adequate design size can be of 11 replicates since:

$$n_{\beta_k} = \left[\frac{t_{\beta_k} s.e._k}{\beta_k}\right]^2 = \left[\frac{1.96 \times 2}{1.2}\right]^2 = 10.67 \approx 11 \tag{13}$$

If the design is segmented into three different subsets consisting of different choice sets, then one would need about 32–33 respondents to achieve 5 per cent significance, assuming that the prior parameter is correct. Such a calculation

[6] Values returned by Equation (11) represent the mathematical minimum sample size requirements, based on the original work by McFadden (1974). Such values represent minimum theoretical sample size requirements, and should be used carefully, particularly given that they are contingent on the priors assumed being correct. Note, however, that in the unlikely event that the priors are exactly correct then the sample size requirement will exactly be the one derived in the formula.

may be made for all K parameters, with the theoretical minimum sample size being the largest value calculated (Bliemer and Rose (2005) proposed that designs that seek to minimise the sample size be termed S-efficient designs). This illustration is informative to clarify the relationship between design and sample size required to achieve significance of β estimates. However, this is obviously a theoretical relationship critically hinging on the exactness of the *a priori* assumption on the β values. In practice such assumption are unlikely to hold perfectly, and the theoretical model supporting the choice of design remains a mere simplification of the real world, so that typically larger sample sizes are necessary than those indicated. How much larger will depend on the empirical case at hand.

4. Design efficiency for prediction and for WTP

In many marketing and transport studies choice experiments are used to derive predictions of choices, and in particular predictions on the effect of changes in the choice attributes. So, other criteria rather than efficiency are used to assess designs when the stated choice exercise has this purpose. Kessels *et al*. (2006) propose the use of G- and V-optimality criteria for the experimental choice context. These criteria measure the variance of prediction, rather than the variance of the taste intensity estimates. In particular, G-optimality relates to the minimisation of the *maximum* prediction variance in the design, while V-optimality relates to the minimisation of the *average* prediction variance.

Finally, of central interest to the literature in non-market valuation and to this paper is the concept of C-optimality, first introduced in the literature by Kanninen (1993a,b). This criterion is specifically suited for minimising the variance of functions of model coefficient estimates, such as willingness to pay. A frequently adopted specification of utility is often specified as a function liner in the parameter of choice attributes, one of which, for valuation studies, is necessarily the cost of the alternative. In this context, it can be shown that the unit WTP for the attribute can be derived as a function of the coefficient attributes:

$$\mathrm{WTP}_k = \frac{\beta_k}{-\beta_{\mathrm{cost}}} \tag{14}$$

This is a highly non-linear function of the coefficient estimates and its variance can be approximated using the delta method.

The ML estimator for β is asymptotically normal, so that given consistency:

$$\sqrt{n}(\beta_{\mathrm{ML}} - \beta) \xrightarrow{D} N(0, \mathrm{Var}(\beta_{\mathrm{ML}})) \tag{15}$$

Take any continuously function differentiable $g(\beta)$. Using the first two terms of a Taylor series approximation to expand it around the estimates one obtains:

$$g(\beta_{ML}) \approx g(\beta) + \nabla g(\beta)'(\beta_{ML} - \beta) \tag{16}$$

Where $\nabla g(\beta)$ is the vector of K first derivatives, the gradient of $g(\cdot)$, and $'$ indicates the transpose.

We can compute the variance of this linear function so that:

$$\text{Var}\left[g(\beta_{ML})\right] \approx \nabla g(\beta)'\text{Var}(\beta_{ML})\nabla g(\beta). \tag{17}$$

With this approximation, all that is now needed is to substitute g(·) with $-\alpha/\beta$. To avoid notational clutter induced by the use of subscripts, we indicate with α the taste intensity of the generic attribute and with β the cost coefficient.

First note that $\alpha/-\beta = -\alpha(\beta)^{-1}$, which makes the use of the product rule to derive the gradient easier:

$$\nabla g(-\alpha\beta^{-1}) = \begin{bmatrix} f' \\ h' \end{bmatrix} = \begin{bmatrix} \dfrac{\partial(-\alpha\beta^{-1})}{\partial\alpha} \\ \dfrac{\partial(-\alpha\beta^{-1})}{\partial\beta} \end{bmatrix} = \begin{bmatrix} -\beta^{-1} \\ \alpha\beta^{-2} \end{bmatrix} \tag{18}$$

So that:

$$\begin{aligned} &\text{Var}\left[g(\beta_{ML})\right] \approx \nabla g(\beta)'\text{Var}(\beta_{ML})\nabla g(\beta) \\ &= \left[-\beta^{-1}\alpha\beta^{-2}\right]\begin{bmatrix} \text{Var}(\alpha) & \text{Cov}(\alpha,\beta) \\ \text{Cov}(\alpha,\beta) & \text{Var}(\beta) \end{bmatrix}\begin{bmatrix} -\beta^{-1} \\ \alpha\beta^{-2} \end{bmatrix} \end{aligned} \tag{19}$$

Multiplying the first row vector by the matrix gives:

$$\left[-\beta^{-1}\text{Var}(\alpha) + \alpha\beta^{-2}\text{Cov}(\alpha,\beta) \quad -\beta^{-1}\text{Cov}(\alpha,\beta) + \alpha\beta^{-2}\text{Var}(\beta)\right] \tag{20}$$

Then, multiplying the resulting row vector by the final column vector gives:

$$\begin{aligned} &-\beta^{-1}\left[-\beta^{-1}\text{Var}(\alpha) + \alpha\beta^{-2}\text{Cov}(\alpha,\beta)\right] \\ &+ \alpha\beta^{-2}\left[-\beta^{-1}\text{Cov}(\alpha,\beta) + \alpha\beta^{-2}\text{Var}(\beta)\right] \rightarrow \text{Var}\left[\frac{\alpha}{-\beta}\right] \\ &\cong \beta^{-2}\left[\text{Var}(\alpha) - 2\alpha\beta^{-1}\text{Cov}(\alpha,\beta) + (\alpha/\beta)^2\text{Var}(\beta)\right] \end{aligned} \tag{21}$$

Thus, the *C*-efficiency criterion relates to the minimisation of such approximation formula for the variance of WTP. One thing to note is that, unlike in the case of CVM in which there is only one WTP to derive, here the variance

relates to an element of $K-1$ WTPs. Furthermore, different attributes may be described in different units. So, for example, with an attribute expressed in miles and one in number of properties affected, the WTP per unit will be referring to different measures. Suppose one takes the sum of the $K-1$ variances, then minimising such a sum may result in an unsatisfactory outcome if the minimum is obtained by diminishing the variance unevenly across WTPs. For example, the minimum may be reached by achieving a very small variance for attribute 1 while leaving the variance for attribute 2 higher than desirable. Equation (13) suggests a potential criterion, which is that of either maximising the minimum t-value for the WTP:

$$x_{sj}^{*} = \underset{x_{sj}}{\arg} \max \left(\min \begin{bmatrix} t_{\mathrm{WTP}_1} \\ \vdots \\ t_{\mathrm{WTP}_{k-1}} \end{bmatrix} \right) \tag{22}$$

or equivalently, that of minimising the number of design replicates necessary to achieve the desired significance level for WTP:

$$x_{sj}^{*} = \underset{x_{sj}}{\arg} \min \left(\max \begin{bmatrix} D_{\mathrm{WTP}_1} \\ \vdots \\ D_{\mathrm{WTP}_{k-1}} \end{bmatrix} \right) \tag{23}$$

To our knowledge neither of these criteria has been used so far in the literature of experiment design for choice studies. We note in passing that all these criteria can be adapted so as to be amenable to a Bayesian prior as discussed in section 3.3.

In conclusion of this review of criteria we emphasise how various criteria are available to evaluate a candidate design and each is particularly suitable to a specific purpose. Of course, when the stated choice exercise has a variety of purposes, then perhaps a weighted combination of selected criteria can be employed to derive the optimal design x_{sj}^{*}. A similar observation can be extended to the final specification. If the data collection is likely to support a variety of specifications, then the AVC matrix may be substituted for an adequate mixture of AVC matrices, one for each specification. However, we do not venture our empirical illustration in this territory, but note that it could constitute fertile ground for further research.

5. What design efficiency measure to report?

For any given choice study, there exist two distinct stages. The first stage relates to the design of the survey instrument and subsequent data collection. The second stage relates to any analysis that is performed on the data collected during the initial stage of the project. Given our discussion above, it is clear that in generating the experimental design during the first stage of a choice study project, the analyst is required to assume the expected values of the

utility parameters of the attributes present within the stated choice task, as well as to the likely model specification to be used during the estimation phase. The efficiency measures discussed above (e.g. the *D*- or *C*-errors) relate only to the statistical efficiency of the design at the generation stage of the study and depend on the assumptions made by the analyst. In other words, they are conditional on the information available at this stage and may or may not be confirmed by the information collected in the data.

With this in mind, we propose a statistical measure comparing how a design is expected to perform, as defined at the design generation phase of the study, against how it actually performed in reality once the data is collected. While such a measure may not necessarily be useful within any one study, if reported over several studies, it may allow researchers to determine how sensitive final model results (in terms of the AVC matrix at least) are to the assumptions that go towards making the design. In turn, this may aid future researchers and practitioners in understanding how much time and effort should be put into generating more statistically efficient designs.

Denoting by the superscript 0 the initial stage priors and with 1 the end of study estimates we recommend future studies report:

$$\frac{F(\hat{\beta}^0, x_{sj})}{F(\hat{\beta}^1, x_{sj}^*)}, \tag{24}$$

where F denotes the particular criterion of interest and the starred design indicates optimisation with respect to the end of study estimates.

Additional criteria might also be reported to understand the relationship between the design employed – which presumably has been derived by optimising according to some valid criterion – and the values that the same design affords with regards to other criteria. So, for example, suppose one has obtained the design x_{sj}^V used in the study by optimising for the V_p criterion of section 4, then it would probably be of interest to contrast this design by using the more common D_p criterion:

$$\frac{D_p(\hat{\beta}^0, x_{sj}^V)}{D_p(\hat{\beta}^0, x_{sj}^D)} \tag{25}$$

A high value of this ratio would illustrate that despite having been derived with a criterion that maximised efficiency in prediction (as *V*-optimality does), such design turns out to perform well relative to the design x_{sj}^D that is optimised for efficiency in coefficient estimates (*D*-efficiency). Different evaluations can also be carried out by using assumed and estimated values of β.

6. Algorithms for design optimisation for efficient designs

We now turn our attention to a brief description of the various algorithms proposed in the literature to search for improvements on or selection from a

basic starting design, which can be, for example, the typical orthogonal fraction of the full factorial. Unfortunately, there does not exist much theoretical guidance as to which method should be employed. We are also not aware of studies that tested which type of design construction method is likely to produce the best results under various circumstances in practice. Several algorithms have been proposed and implemented within the literature to systematically search the various arrangements of attribute levels and identify efficient designs. These algorithms operate mostly by systematically operating swaps across the rows and columns of the matrix x_{sj}. Typically, algorithms fall into one of two categories; row-based and column-based algorithms.

In *row-based algorithms*, a large number of choice sets are first generated from which those to be used in the survey are selected. Typically, the choice sets are drawn from a full factorial design, although in many instances the full factorial will be too large (even with today's computing power) and fractional factorials may be generated instead. This is precisely what the most widely used row-based algorithm, the *Modified Federov algorithm* (Cook and Nachtsheim 1980), does. The algorithm randomly draws s choice sets from either a full factorial or fractional factorial design, and computes the D-error of each random selection. The combination of choice sets that produce the lowest D-error is retained as the most efficient design. The algorithm is terminated either manually by the researcher, when some stopping criteria is achieved (e.g. no improvement in the D-error is achieved for 30 min) or when all possible choice set combinations have been explored. Row-based algorithms have the advantage of being able to reject poor choice set candidates at the initial stage (e.g. choice sets in which the attributes of one or more alternatives are dominated or where a particular combination of attributes realistically cannot exist), and as such, these choice sets will never appear in the final survey. Nevertheless, row-based algorithms generally find it difficult to maintain attribute level balance (where each attribute level appears an equal number of times over the design).

Column-based algorithms on the other hand, begin by randomly generating a design and then systematically change the levels within each column (representing an attribute in the survey) of the design. While it is difficult to reject poor choice sets using column-based algorithms, such algorithms typically are able to maintain attribute level balance, particularly if the initially generated design has such a property. In general, column-based algorithms offer more flexibility and are generally easier to use when dealing with designs with many choice situations, but in some cases (e.g. for unlabelled choice experiments and for specific designs such as those where certain attribute level combinations are forbidden) row-based algorithms may be more suitable.

Rather than relying solely on row- or column-based algorithms, some authors suggest using combinations of both. Huber and Zwerina (1996) implemented the RSC algorithm (*Relabelling*, *Swapping* and *Cycling*), which remains the

most widely used algorithm today. The RSC algorithm alternates between *relabelling* (column-based), *swapping* (column-based), and *cycling* (row-based) over many iterations. During the *relabelling* phase, all occurrences for two or more attribute levels within a column of the design are switched (e.g. if attribute levels 1 and 4 are relabelled then the column containing the sequence of levels {1,3,4,2,4,1,3,2} would become {4,3,1,2,1,4,3,2}). The *swapping* phase of the algorithm is similar to that of relabelling, however, only a few of the attribute levels are changed within the column (e.g. swapping the first and third values in {1,3,4,2,4,1,3,2} would yield {4,3,1,2,4,1,3,2}). The *cycling* phase of the algorithm is row-based, where the attribute levels are switched (similar to relabelling but now across rows, not down columns) within choice sets, one choice set at a time. The algorithm will generally try a number of iterations of either *relabelling*, *swapping* or *cycling*, before switching to another phase (typically randomly). Note that not all phases have to be used with various combinations of RSC being possible.

7. The impact of scale on willingness to pay

One consideration must be made at this stage about the scale parameter λ, of equation (1), which is often a neglected issue in *D*-efficient designs. This is particularly relevant when the focus is on WTP estimation and when a status-quo constant (or any alternative-specific constant) is expected to be part of the utility function, as is often the case in non-market valuation studies. WTP computations are one-to-many mappings of the β vector. In fact, infinite pairs of β_{-p} (non-price coefficients) and β_p produce the same vector of WTP values. Suppose, the values of β are as assumed above. Scaling them all by any positive constant produces the same WTP estimates. So, implicit in the assumption of values for β there is an assumption of the scale coefficient.

When – instead – utility includes an alternative-specific constant of some sort, scaling the vector β by any amount has an effect on the utility differences across alternatives, which are not scaled by the same constant. So, depending on the assumed scale parameter of the Gumbel error, the same WTP vector can be associated with large or small utility differences with the status-quo, and hence different choice probabilities. Table 1 illustrates this case in which the levels of the attributes in the status-quo (SQ) choice are assumed to be the baseline (equal to zero) and hence the levels in the designed alternatives 1 and 2 are expressed as differences from those in the SQ.

This is, of course, a corollary to the fact that with a high scale (small error variance) the choice probabilities become deterministic. However, it highlights how important an adequate specification of the error scale is to the evaluation of the design in the presence of alternative-specific constants. For a given scale though, the criteria of different designs can be compared. We hence now turn to a comparison of designs generated under the assumption of a multinomial logit specification for a given case study.

Table 1 Demonstration of impact of scaling on model outcomes

$\lambda = 1$								
β	1	1	2	3	–1	$\beta' x_j$	$\Delta V_{j\text{–sq}}$	Pr(j)
x_1	0	1	2	2	2	9	8	0.952
x_2	0	2	2	1	3	6	5	0.047
Sq	1	0	0	0	0	1	0	0.000
WTP	1	1	2	3	–1	–	–	–
$\lambda = 0.5$								
β	0.5	0.5	1	1.5	–0.5	$\beta' x_j$	$\Delta V_{j\text{–sq}}$	Pr(j)
x_1	0	1	2	2	2	4.5	4	0.806
x_2	0	2	2	1	3	3	2.5	0.180
Sq	1	0	0	0	0	0.5	0	0.015
WTP	1	1	2	3	–1	–	–	–
$\lambda = 0.2$								
β	0.2	0.2	0.4	0.6	–0.2	$\beta' x_j$	$\Delta V_{j\text{–sq}}$	Pr(j)
x_1	0	1	2	2	2	1.8	1.6	0.571
x_2	0	2	2	1	3	1.2	1	0.313
Sq	1	0	0	0	0	0.2	0	0.115
WTP	1	1	2	3	–1	–	–	–

8. Case study

8.1 The case study setting

This case study is devised to illustrate the considerations a researcher can make when engaged in developing a 'typical' non-market valuation study. A recent review on the design solutions used in published non-market valuation studies (Ferrini and Scarpa 2007) suggests that a common set up is an unlabelled design based on a choice task involving the indication of the favourite alternative among three. Two of these have levels and attributes developed on the basis of a design, while the third represents the status-quo (see Breffle and Rowe (2002) for a discussion of the inclusion of the status-quo alternatives in non-market valuation studies and Scarpa *et al*. (2005) for some econometric insights). We hence adopt this framework, but caution the reader that generalising the results from this case study to other contexts might well be unwarranted.

Most published studies investigate a range of 3–6 choice attributes plus the cost of the package to the respondent. We hence present results of a design with three attributes plus price and a status-quo constant. We postulate that the analyst is able to define some *a priori* beliefs on the values of the β vector that can be adequately formalised. We assume that since much of the literature reports positive status-quo effects, the element of β relating to the status-quo is assumed to be positive and equal to unity. The price effect is of course negative and also equal to one. The three attributes differentiating the alternatives are assumed to be expressed as positive effects on utility and orderable in terms of a gradient one, two, and three. Thus, the utility functions used in generating the designs for the case study may be summarised as follows.

$$\begin{aligned} V_1 &= 1x_{11} + 2x_{21} + 3x_{31} - 1x_{41}, \\ V_2 &= 1x_{12} + 2x_{22} + 3x_{32} - 1x_{42}, \\ V_{sq} &= 1. \end{aligned} \tag{26}$$

While one can very frequently express attributes in a way that can be generally expected to be perceived and evaluated by respondents as having a specific directional (positive or negative) effect on utility, the cardinal scaling is arguably the strongest *a priori assumption.* However, this assumption can be relaxed by assuming a distributional form with overlapping densities, as we will see later.

The size of the design is of 20 choice sets, and the design attributes can all take four values (0,1,2,3) except price which can take five levels (these are 0,1,2,3,4). A size of 20 is not unusual and can be shared out across five, four or two respondents to obtain a balanced panel of, respectively, four, five and ten choices per respondent. In non-market valuation studies it is frequently found that the number of levels used for the price attribute is larger than those used for non-price attributes.

8.2 Exploration of design procedure

Fifteen designs are generated and compared across a range of criteria. In order to demonstrate why it is important to use experimental designs for stated preference studies, the first two designs we report were constructed using a purely random allocation of the attribute levels to the design. In generating the first design, we do not assume attribute level balance (i.e. each level of an attribute may appear an uneven number of times over the 20 choice sets), whereas for the second design, attribute level balance was enforced as a design criteria. All remaining designs also assume attribute level balance. Unlike Designs 1 and 2, Designs 3 to 5 and 9 to 15 were constructed using the RSC algorithm (see Section 5) assuming (different) optimisation criterion. Design 6 was constructed in a manner for which the RSC algorithm was not appropriate and hence only swapping was used. Designs 7 and 8 are orthogonal designs, for which the RSC algorithm is also inappropriate.

Designs 3, 4 and 5 represent designs constructing using the *D*-efficiency criterion given as Equation (5), and they illustrate the effect of varying the scale parameter in this context, as discussed in Section 7. In generating Design 3, we assumed as prior parameter estimates, the values discussed above. In Design 4 we double the magnitude of the prior parameter estimates, whereas Design 5 halves the magnitudes. The sixth design was constructed also using the *D*-efficiency criterion. However, many restrictions were placed on such design. Specifically, the design was generated so that the attribute levels for one of the non-status-quo alternatives are always lower than that of the other non-status-quo alternative. Given that higher levels for the non-price attributes are assumed to be more preferred (i.e. the prior parameters assumed were all positive for these attributes) while higher prices

are obviously less preferred (i.e. a negative prior parameter) this design forces respondents to trade (simultaneously) the non-price attributes with price within each choice set. Such a constraint is designed to ensure that some form of trading always takes place in choice tasks. However, we note that strictly speaking one cannot assume that generating a design in this manner will avoid dominance in terms of preferences.[7]

Designs 7 and 8 are orthogonal fractional factorial designs. In constructing the designs, no orthogonal design could be found that allowed for zero correlations both within and between the attributes of alternatives. So, a sequential design process was employed (see Louviere *et al.* 2000). This process involves first constructing an orthogonal design for alternative 1, and then using the same design to construct alternative 2. The process ensures that the designs are orthogonal in the attributes within alternatives, but not between alternatives. Given that the experiment is assumed to be unlabelled, the between alternative correlations are not of concern and hence the design process is appropriate. While maintaining the (within alternative) orthogonality constraint, the *D*-efficient criteria was also applied to Design 8.

Designs 9, 10 and 11 are non-orthogonal designs generated to minimise, respectively, *A*-, *S*- and *B*-criteria, of Equations (6), (12), and (7), respectively. The remaining designs are generated in such a way as to minimise the sum of the *C*-efficiency measures in Equation (22). Designs 12 and 14 consider only the variances of the WTP values for the design attributes, whereas Designs 13 and 15 also consider the variance of the WTP for the status-quo constant. To illustrate the flexibility afforded by applying the *C*-criterion we use different weights for the variances of the WTPs of different attributes when generating the last two designs. So that the criterion employed is the minimisation of the weighted sum of the variance components of the attribute WTPs. This flexibility may be important in practice when the object of a stated preference study is to specifically calculate the WTP for a subset of the design attributes. The full set may include attributes considered important within the preference space of the respondent, but irrelevant from the viewpoint of WTP estimation. Alternatively, the absolute magnitudes of the WTP outcomes may also guide whether weighting should be applied, for example, whether it is to be expressed in dollars or cents. For the present study, in constructing Design 14, attribute 1 is assigned the largest value of 0.4 because it is the one with

[7] Dominance implies that all respondents acting rationally will always select one alternative over all others present. Design 6 ensures that respondents will be faced with a comparison between a lower '*quality*' lower price alternative and a higher *quality* higher price alternative, but says nothing about the probability that one of the alternatives will be chosen. To establish whether an alternative is dominated or not, the analyst would need to calculate the choice probabilities (which are function of the design attributes and (prior) parameters). Once the choice probabilities are determined, the analyst would need to establish some rule as to what constitutes a dominated alternative based on the expected choice probabilities (e.g. if the probability is less than 0.1).

lowest absolute WTP. As such, more precision (efficiency) is needed for this attribute compared to the others to obtain a WTP estimate different from zero. For similar reasons attribute 2 is assigned a value of 0.35, and Attribute 3 of 0.25. The status-quo constant is ignored in this design, and hence has a weight of zero. A similar weighting procedure is applied in generating Design 15, with weights of 0.4, 0.3, 0.2 and 0.1 being applied to each of the design attributes and status-quo constant, respectively.

All designs were generated using either Microsoft Excel or a program called Ngene. Visual Basic macros were constructed in Microsoft Excel with the AVC matrices of the designs constructed using matrix algebra manipulation formulas that are standard within Microsoft Excel. Ngene is a design generation program currently under development by Econometric Software.

8.3 Design outcomes

For each of the 15 designs we generated Tables 2 and 3 present various measures of the design criteria discussed. For each measure, excluding the *B*-statistic, values are presented based on computations including and excluding the status-quo constant. As would be expected, the two random designs perform very poorly on each design criterion measure presented in the table. This outcome, however, is based on random chance, and different results might have been obtained if a different random allocation of the attribute levels were considered. The design obtained by minimising the *D*-error (Design 3) appears to perform very well on all criteria except for the *B*-statistics. According to the *S*-error for the design, a minimum of seven replications of the design (representing 140 choice tasks) are required for all parameters, including the status-quo constant to be statistically significant at the 1.96 level. Of course, this number assumes that the prior parameter used is correct, hence, this represents only the theoretical minimum number of design replications that should be collected.

Designs 4 and 5 illustrate the impact of assuming different prior values for the scale parameter λ, when generating the design. Contradictory results are produced when doubling and halving λ. The importance of accounting for scale size has eloquently been described by Swait and Louviere (1993), and the reader is reminded here that higher scale implies smaller variance and that as scale increases the choice becomes gradually more deterministic. In our context halving the priors produces superior *D*- and *A*-error results, but dramatically worsens results in terms of WTP and sample size requirements when compared to a doubling of the scale parameter. These results might be perceived as counter-intuitive, as one would expect that doubling the assumed scale of the error (Design 4) and hence increasing the precision should lead to a higher efficiency. Instead, one observes the opposite. Increasing scale decreases the information content of the design for β while it increases it for the attribute WTPs. One possible cause for this might be that in generating the design, the attribute levels used are the same as those used for Design 3,

Table 2 Efficiency level outcomes for Designs 1–15

Design	Effect	*D*-error		*A*-error		*C*-error		Weighted *C*-error		*S*-error		*B*-error
		Without Constant	With Constant	Without Constant	With Constant	Without Constant	With Constant	Without Constant	With Constant	Without Constant	With Constant	
1	Base Design (random – unbalanced)	0.998	2.136	2.306	17.170	10.076	111.894	–	–	10.076	294.379	0.06%
2	Base Design (random – balanced)	0.920	1.398	2.202	4.847	8.630	22.677	–	–	9.218	59.270	1.67%
3	*D*-error	0.120	0.189	0.909	1.052	2.030	0.519	–	–	1.001	6.238	10.03%
4	Scale up ($\beta \times 2$)	0.198	0.290	3.895	3.612	0.176	0.656	–	–	1.015	9.529	7.77%
5	Scale down ($\beta \times 0.5$)	0.076	0.126	0.200	0.448	2.034	7.955	–	–	1.396	22.070	21.96%
6	Constrained trade-off	2.768	2.436	8.629	7.586	29.682	35.690	–	–	10.896	13.104	3.06%
7	Random orthogonal	1.828	2.146	11.828	4.326	24.393	160.706	–	–	8.786	15.368	5.10%
8	Efficient orthogonal	0.334	0.464	1.318	1.024	1.643	4.160	–	–	1.300	9.592	5.25%
9	*A*-error	0.212	0.283	0.526	0.653	1.230	2.503	–	–	0.943	4.456	10.25%
10	*S*-efficient	0.373	0.408	1.589	1.407	2.594	1.487	–	–	2.189	2.602	23.92%
11	*B*-error	0.384	0.419	1.879	2.057	4.391	1.926	–	–	10.634	3.778	41.20%
12	*C*-error (attributes only)	0.153	0.281	2.984	4.282	0.455	6.456	–	–	3.293	36.386	7.53%
13	*C*-error (attributes + SQ)	0.206	0.262	3.185	2.838	0.551	1.454	–	–	3.454	5.585	21.42%
14	Weighted *C*-error (attributes only)	0.244	0.302	5.821	5.120	0.540	1.496	0.666	0.666	6.347	8.902	28.16%
15	Weighted *C*-error (attributes + SQ)	0.183	0.251	3.043	2.778	0.501	1.601	0.526	0.966	3.527	6.602	17.13%

Table 3 *T*-ratio (assuming a single design replication) and minimum design replication requirements by attribute for Designs 1–15

	β		WTP		β		WTP		β		WTP	
	|*t*-values|	*n*	|*t*-values|	*n*	|*t*-values|	*n*	|*t*-values|	*n*	|*t*-values|	*n*	|*t*-values|	*n*
	Design 1 Random allocation (unbalanced)				Design 2 Random allocation (balanced)				Design 3 *D*-efficient			
Constant	0.255	59.270	0.267	53.962	0.114	294.379	0.106	340.096	0.785	6.238	0.814	5.804
$\beta 1$	0.646	9.218	0.858	5.219	1.113	3.103	0.737	7.068	1.959	1.001	3.589	0.298
$\beta 2$	1.427	1.886	1.347	2.118	1.681	1.359	0.647	9.173	2.000	0.960	5.016	0.153
$\beta 3$	1.712	1.311	1.333	2.163	1.434	1.868	0.867	5.111	2.061	0.904	5.642	0.121
$\beta 4$	0.854	5.268	n.a.	n.a.	0.617	10.076	n.a.	n.a.	1.977	0.983	n.a.	n.a.
	Design 4 $\beta \times 2$				Design 5 $\beta \times 0.5$				Design 6 Trade-off constrained			
Constant	0.635	9.529	0.721	7.385	0.417	22.070	0.411	22.749	0.541	13.104	0.408	23.079
$\beta 1$	1.951	1.010	7.294	0.072	1.659	1.396	1.744	1.263	0.612	10.243	0.567	11.961
$\beta 2$	1.961	0.999	8.223	0.057	2.194	0.798	2.649	0.548	0.594	10.896	0.578	11.517
$\beta 3$	1.966	0.994	9.596	0.042	2.279	0.740	2.816	0.484	0.689	8.097	0.786	6.222
$\beta 4$	1.945	1.015	n.a.	n.a.	1.893	1.072	n.a.	n.a.	0.807	5.904	n.a.	n.a.
	Design 7 Orthogonal				Design 8 Orthogonal efficient				Design 9 *A*-efficient			
Constant	0.155	160.706	0.139	199.680	0.633	9.592	0.630	9.667	0.928	4.456	0.886	4.891
$\beta 1$	0.397	24.393	0.311	39.792	1.719	1.300	2.074	0.894	2.018	0.943	2.186	0.804
$\beta 2$	1.347	2.116	1.113	3.103	1.852	1.121	2.554	0.589	2.632	0.554	3.171	0.382
$\beta 3$	1.086	3.258	1.336	2.153	1.959	1.001	3.360	0.340	2.868	0.467	3.802	0.266
$\beta 4$	0.945	4.304	n.a.	n.a.	2.020	0.941	n.a.	n.a.	2.310	0.720	n.a.	n.a.
	Design 10 *S*-efficient				Design 11 *B*-efficient				Design 12 C_p-efficient attributes only			
Constant	1.215	2.602	0.950	4.256	0.601	10.634	0.637	9.470	0.325	36.386	0.408	23.051
$\beta 1$	1.325	2.189	1.921	1.041	1.008	3.778	1.280	2.346	1.080	3.293	3.881	0.255
$\beta 2$	1.632	1.443	2.957	0.439	1.463	1.796	3.200	0.375	1.109	3.122	5.323	0.136
$\beta 3$	1.552	1.594	3.446	0.324	1.474	1.769	3.120	0.395	1.133	2.995	6.025	0.106
$\beta 4$	1.347	2.117	n.a.	n.a.	1.390	1.989	n.a.	n.a.	1.109	3.125	n.a.	n.a.
	Design 13 C_p-efficient attributes + sq				Design 14 Weighted C_p-efficient attributes only				Design 15 Weighted C_p-efficient attributes + sq			
Constant	0.829	5.585	1.052	3.470	0.657	8.902	1.023	3.671	0.763	6.602	0.954	4.223
$\beta 1$	1.064	3.391	3.333	0.346	0.778	6.347	3.478	0.318	1.044	3.527	3.525	0.309
$\beta 2$	1.055	3.454	4.714	0.173	0.800	5.995	4.563	0.185	1.105	3.145	5.127	0.146
$\beta 3$	1.100	3.175	5.662	0.120	0.807	5.904	5.826	0.113	1.122	3.051	5.786	0.115
$\beta 4$	1.103	3.156	n.a.	n.a.	0.801	5.987	n.a.	n.a.	1.097	3.192	n.a.	n.a.

and since it is differences in utility that matter most, the utility differences observed with a scaled up set of β are larger and induce large variations in choice probabilities, at the expense of design balance and information content. The opposite effect is at play for Design 5, where the scale of the priors is half that of Design 3.

Ignoring the designs that were randomly generated and where the parameter priors have been re-scaled, Design 6 performs quite poorly based on all criteria when compared to the other designs. This is because the trade-off constraint, while attempting to conform to some analyst-imposed behavioural heuristic, fails to consider the statistical requirements that improve the statistical efficiency of experimental designs. In particular, the AVC matrix of a design, from which all efficiency measures are derived (save for the *B*-error measure), is the inverse of the second derivatives of the log-likelihood function for the design. As such, the AVC matrix is intrinsically related to the choice probabilities that the design is likely to produce (given prior parameter estimates). In setting up the (behavioural) constraint, the expected choice probabilities for the design are also constrained, which in turn impacts on the design AVC matrix and its efficiency. So, such design strategy, while behaviourally attractive, is likely to produce poor outcomes in terms of model efficiency.

Design 7 represents the currently predominant method used for generating stated choice experimental designs; the generation of an orthogonal design (see Louviere *et al*. (2000) Ch5 or Hensher *et al*. (2005) Ch5). However, as shown here, the use of orthogonal designs tends to produce less than optimal outcomes in terms of expected model results, requiring larger sample sizes to retrieve statistically significant parameter estimates than other non-orthogonal designs. Design 8 represents an improvement on Design 7 and it is derived by employing an algorithm that minimises the *D*-error of the design while maintaining orthogonality.[8] Even so, the imposition of orthogonality represents a constraint on the efficiency of stated choice designs, for the exact same reasons as given for the poor performance of Design 6. That is, the imposition of orthogonality only relates to the correlation structure of the design, but says

[8] For problems where it is possible to generate an orthogonal design, it may be possible to locate more than one such design. When constructed from first principles, orthogonal fractional factorial designs are typically generated by confounding higher order interaction effects. Confounding different interaction effects may result in different orthogonal fractional factorial designs. As such, it is often possible to generate more than on orthogonal fractional factorial design for a given problem. Nevertheless, researchers typically only generate one such design and fail to consider that other possible fractional factorial designs can be generated. Further, once an orthogonal fractional design is located, it is often possible to switch the levels of the design in such a way to maintain orthogonality but produce a different orthogonal design. This is similar to the relabelling phase of the RSC algorithm. In the current case study, design 7 is a randomly selected orthogonal fractional factorial design, whereas design 8 is an orthogonal design that is chosen such that it produces the lowest possible *D*-error while maintaining the orthogonality constraint.

nothing of the choice probabilities and hence the AVC matrix that the design will produce.[9]

Designs 9 to 11 were constructed so as to optimise the measures for *A*-, *S*- and *B*-criteria, respectively. In each case, the designs produce the lowest (highest for the *B*-criterion design) values for the respective criterion for which the design was optimised. These designs appear to perform very similarly on all other criteria, however, Design 11 – which optimised the *B*-criterion – appears to require the largest minimum number of design replications to retrieve statistically significant parameters and WTP values. This finding is consistent with Sándor and Wedel (2001, 2002, 2005) and Kanninen (2002) who demonstrated that complete utility balance, as explored by Huber and Zwerina (1996), will result in suboptimal designs.

Our last group of comparisons are made across designs obtained by using various specifications of *C*-efficiency as the optimisation criteria. These designs perform well compared to most other designs, however, several issues arise which require further discussion. Firstly, assuming the priors have been correctly specified, the theoretical minimum number of design replications required for Design 12 is 37 (i.e. 740 choice observations) for all parameters to be statistically significant as per Equation (11). Table 3, demonstrates the asymptotic *t*-ratios for each attribute and WTP for each design, as well as the number of design replications required in order for the asymptotic *t*-ratios to be greater than 1.96. An examination of this table for Design 12 shows that the requirement for 37 replications of the design is a result of the status-quo constant, which was not considered when generating the design. As such, it is questionable as to whether one would consider 37 replications or the next highest multiple of four replications to be the minimum.

A second observation relates to the use of the *C*-efficiency criteria as expressed previously. The *C*-efficiency criteria, as implemented here, relates only to the variances of the ratios of two parameters, and not the variances of the parameters themselves. While there exists a relationship between the two, the additional non-variance terms contained within Equation (21) may compensate for larger parameter variances when minimising the equation. As such, it may be possible to minimise the variance of the ratio of the two parameters while obtaining a relatively large variance for one or more of the parameters themselves. This has implications when calculating the WTP for that attribute and it is clearly demonstrated in Table 3. Consider for example, Design 13. For the status-quo constant term to achieve an asymptotic *t*-ratio of 1.96, at least six (rounding up from 5.585) design replications are required (120 choice observations), whereas only four (rounding up from 3.470)

[9] Strictly speaking this statement is false. An orthogonal design will be optimal when all parameter priors are assumed to be zero (that is, irrelevant in the decision process). As such, orthogonal designs will only require the smallest possible design replications relative to all other designs when one is willing to assume that the attributes in the design do not play a role in the observed choices. This is obviously contrary to the spirit of most investigations.

replications are required (80 choice observations) for the WTP for the status-quo constant term to achieve statistical significance. Given that the WTP for an attribute should only be calculated if the individual parameters are statistically significant, the higher value of the two should be used (i.e. six design replications). A search through Table 3 reveals that Designs 9 (*A*-efficiency) and 10 (*S*-efficiency), while requiring a larger number of design replications for all WTP values to become statistically significant, would require only three and five design replications (i.e. 60 and 100 choice observations) respectively for all parameter and WTP values to be statistically efficient. As such, these designs would be preferred based on these criteria.

While we do not implement it here, it should be possible to create a new optimisation criterion, similar to the *S*-efficiency measure that minimises the largest sample size required for the ratios of two parameters (the WTP) to be statistically significant. Indeed, one could combine this with the current *S*-efficiency measure for the utility parameters, and jointly minimise both.

9. Conclusion and direction of further research

The use of stated preference methods has become increasingly accepted in the policy arena as a way to investigate non-market values worldwide. Yet, choice modelling has not been subject to the degree of investigation and scrutiny dedicated to contingent valuation in the non-market valuation literature. With particular regards to the topic of experimental design tailored to the specific needs of non-market valuation practitioners the literature is still scarce. This study had the objective of bringing together a number of considerations and design statistics that the practitioner could find of interest. In particular, the principles outlined here can be adopted in the evaluation of choice model designs predicated under different assumptions from the one used for convenience here as the main example.

C-efficiency, for example, is a criterion for design evaluation that although proposed over 15 years ago, is still rarely used. Sample size determination, as we explained here can be theoretically linked to design properties, and can itself be used as a criterion for design search. Importantly, we suggest alternative ways of reporting design statistics in applied studies that go beyond the frequently used percent efficiency criterion originally proposed for multivariate linear regression studies explaining treatment effects in agricultural experiments. We show how this criterion is irrelevant and a bad proxy for *C*-efficiency, which is what matters when the focus is WTP estimation. Nonmarket studies are often plagued by limited budgets, which make design efficiency a prominent feature. Studies focusing on WTP estimation can benefit from the use of the *C*-efficiency criterion because it is tailored to WTP estimates. Designs maximising *C*-efficiency are shown here to outperform those derived according to the more common *D*-efficiency criterion.

In this paper, we have also discussed several other possible criteria on which the efficiency of various designs can be judged. Questions persist as to

which design criteria are most appropriate. While we have argued that some criteria are inappropriate for stated choice studies based on efficient logit analysis, the issue remains as to which criteria from the remaining set of possible measures should be employed by researchers. Unfortunately, there exists no general answer to this question, but the right criteria that are case dependent need to be selected in the light of the objectives of each study. Thus, if one wishes to produce model results that minimise both the standard errors and covariances of the parameter estimates, then the *D*-error criterion should be employed. If, on the other hand, the researcher is limited in budget and wishes to minimise sample size requirements, then the *S*-error is the appropriate measure to use. If the objective of the study is to examine WTP for various attributes, then the *C*-error measure is the appropriate statistic to use during the design generation process. Of course, as we have done in the case study, it is possible to compute more than one criterion for any given design. It is also theoretically possible to optimise a design based on some form of weighted efficiency measure, taking into account more than one objective (e.g. both *S*- and *C*-errors might simultaneously be considered). Future research should investigate options of this kind given that many studies might embrace multiple objectives.

What is certain is that probability balance (*B*-efficiency) should not be used, and that orthogonal designs should be avoided when one seeks efficiency in a context of logit specifications. While orthogonal designs are commonly used within the literature, we have argued that such designs are inefficient for the types of non-linear models used in stated choice studies and can usually be improved upon in terms of the robustness of the parameter estimates. Nevertheless, as demonstrated here, when orthogonal designs are used, it might be possible to generate several orthogonal designs and select the one, for example, that is best suited for a given criterion. For example, the analyst may select an orthogonal design that is most likely to result in the smallest standard errors for the ratio of two parameters, assuming that WTP estimation is what is of interest.

Of course, in sufficiently large samples, statistically significant parameter estimates should be retrieved independent of the final design employed. Indeed, asymptotically, all designs enabling identification of the effects of interest, whether efficient or simply consisting of randomly assigned attribute level combinations should reproduce the true population level parameters. Thus, if the budget for a study is such that the final sample size is not a concern, then the analyst need not worry about the experimental design. In principle, one could simply allocate randomly attribute levels to choice sets and these across respondents. In smaller sample sizes however, the design employed can play a significant role in whether the final model results are statistically efficient or not.

The question then arises, however, as to what to do if there is sufficient budget and several possible designs are expected to produce statistically efficient parameter estimates. For example, what if there exists a budget for data

collection based on 300 respondents and it is calculated that an orthogonal design will require 280 respondents while an efficient design predicated on *a priori* assumptions will require only 200 respondents. Does it matter, which design is used? The answer is probably dependent on the confidence the researcher has on the assumptions required to develop an efficient design. If a good level of confidence is available, we would argue that at least to start with the efficient design should be used with a sample size of 200 respondents. Once the responses based on such design are made available, one can combine the new information from the data with the existing prior in a Bayesian sequential design (e.g. Scarpa *et al*. 2007). Depending on how accurate the initial prior turns out to be, one might stop the data collection and save the budget necessary for the additional 100 respondents. Alternatively, if the initially assumed design is in disagreement with the data, one may use the information collected with the data to update the design before producing an improved design for the last 100 observations, so as to reach the desired final efficiency with an improved understanding of the population parameters. We would suggest that committing money from the beginning to all 300 respondents when the same results could be achieved based on 200 respondents is wasteful of finite resources, and precludes the researcher the opportunity to improve on the initial design at a later stage. Secondly, we remind the reader that sample size calculations represent theoretical minima based on the assumptions made at the time the design is generated. As such, if the assumptions do not hold in practice, the actual sample sizes required might be higher for both designs. As such, if given a sample of 200 respondents it is found that the parameter estimates from the efficient design are unstable (that is still changing significantly with each additional respondent added) and the parameters are statistically not significant, then there remains scope to increase the sample size. If however, the orthogonal design is used, this option might be precluded because of its lower efficiency. As such, we would argue that efficient designs (independent of the criteria used to define efficiency) should always be preferred to start with, even if larger sample sizes are affordable.

We have intentionally neglected several important considerations related to the behavioural efficiency of the design, concentrating our focus on the statistical efficiency and the comparison of different criteria to practically measure it. Future research should focus on respondent efficiency as well. Although perhaps the current level of knowledge on how respondents process the information provided in choice tasks is still insufficient to derive efficiency measures to evaluate behavioural efficiency, this knowledge gap is filling quickly. For example, extensive research has been conducted on the impact upon behavioural responses given various design dimensions. For example, the number of alternatives within the task (Hensher *et al*. 2001), the number of attributes (Pullman *et al*. 1999), the number of attributes and alternatives (DeShazo and Fermo 2002; Arentze *et al*. 2003), the impact of attribute level range upon response (Cooke and Mellers 1995; Ohler *et al*. 2000; Verlegh

et al. 2002) and the number of choice profiles shown to respondents (Brazell and Louviere 1998) have all been examined. More recently, Hensher (2004, 2006a,b) and Caussade *et al*. (2005) examined all of the above effects simultaneously. Nevertheless, an examination of the combination of the design and respondent efficiency remains to date, ever elusive.

One final note is required. While we recommend that choice studies report the ratio of the design criteria to that of the final estimated model, we do not report this statistic here. The reason for this is that in this paper we have concentrated solely on the design generation phase in the case study. No actual data has been collected, with the results reported derived from analytical simulations. As such, the actual performance of the designs in practice cannot be tested. This constitutes an obvious limitation of the current study. However, we note from a review of the literature that all such results on efficient designs are constrained by theory. Research efforts are currently being undertaken by several researchers where various designs are being used and tested in terms of their performance in practice.

References

Alberini, A. (1995). Optimal designs for discrete choice contingent valuation surveys: single-bound, double bound and bivariate models, *Journal of Environmental Economics and Management* 28, 287–306.

Arentze, T., Borgers, A., Timmermans, H. and DelMistro, R. (2003). Transport stated choice responses: effects of task complexity, presentation format and literacy, *Transportation Research Part B* 39, 229–244.

Bliemer, M.C.J. and Rose, J.M. (2005). Efficiency and Sample Size Requirements for Stated Choice Studies, working paper: ITLS-WP-05-08.

Bliemer, M.C.J. and Rose, J.M. (2006). Designing Stated Choice Experiments: State-of-the-art, Paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan.

Bliemer, M.C.J., Rose, J.M. and Hensher, D.A. (in press). Constructing Efficient Stated Choice Experiments Allowing for Differences in Error Variances Across Subsets of Alternatives, *Transportation Research Part B*.

Brazell, J.D. and Louviere, J.J. (1998). Length effects in conjoint choice experiments and surveys: an explanation based on cumulative cognitive burden. Working Paper, Department of Marketing, The University of Sydney, July.

Breffle, W.S. and Rowe, R.D. (2002). Comparing choice question formats for evaluating natural resource tradeoffs, *Land Economics* 78, 298–314.

Burgess, L. and Street, D.J. (2005). Optimal designs for choice experiments with asymmetric attributes, *Journal of Statistical Planning and Inference* 134, 288–301.

Caussade, S., Otúzar, J. de D., Rizzi, L.I. and Hensher, D.A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates, *Transportation Research B* 39, 621–640.

Cook, R.D. and Nachtsheim, C.J. (1980). A comparison of algorithms for constructing exact *D*-optimal designs, *Techometrics* 22, 315–324.

Cooke, A.D. and Mellers, B.A. (1995). Attribute range and response range: limits of compatibility in multiattribute judgment, *Organizational Behavior and Human Decision Processes* 63, 187–194.

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

Ferrini, S. and Scarpa, R. (2007). Designs with a-priori information for nonmarket valuation with choice-experiments: a Monte Carlo study, *Journal of Environmental Economics and Management* 53, 342–363.

Hensher, D.A. (2004). Accounting for stated choice design dimensionality in willingness to pay for travel time savings, *Journal of Transport Economics and Policy* 38, 425–446.

Hensher, D.A. (2006a). Revealing differences in behavioural response due to the dimensionality of stated choice designs: an initial assessment, *Environmental and Resource Economics* 34, 7–44.

Hensher, D.A. (2006b). How do respondents process stated choice experiments? Attribute consideration under varying information load, *Journal of Applied Econometrics* 21, 861–878.

Hensher, D.A., Stopher, P.R. and Louviere, J.J. (2001). An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: an airline choice application, *Journal of Air Transport Management* 7, 373–379.

Hensher, D.A., Rose, J.M. and Greene, W.H. (2005). *Applied Choice Analysis: A Primer,* Cambridge University Press, Cambridge.

Huber, J. and Zwerina, K. (1996). The importance of utility balance in efficient choice designs, *Journal of Marketing Research* 33, 307–317.

Kanninen, B.J. (1993a). Optimal experimental design for double bounded dichotomous choice contingent valuation, *Land Economics* 69, 138–146.

Kanninen, B.J. (1993b). Design of sequential experiments for CV studies, *Journal of Environmental Economics and Management* 25, 1–11.

Kanninen, B.J. (2002). Optimal designs for multinomial choice experiment, *Journal of Marketing Research* 39, 214–227.

Kessels, R., Goos, P. and Vandebroek, M. (2004). *Comparing Algorithms and Criteria for Designing Bayesian Conjoint Choice Experiments*. Working paper, Department of Applied Economics, Katholieke Universiteit Leuven, Belgium.

Kessels, R., Goos, P. and Vandebroek, R. (2006). A comparison of criteria to design efficient choice experiments, *Journal of Marketing Research* 43, 409–419.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000). *Stated Choice Methods – Analysis and Application*. Cambridge University Press, UK.

Lusk, J.L. and Norwood, F.B. (2005). Effect of experimental design on choice-based conjoint valuation estimates, *American Journal of Agricultural Economics* 87, 771–785.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour, in Zarembka, P. (ed.), *Frontiers in Econometrics*. Academic Press, New York, pp. 105–142.

Ohler, T., Li, A., Louviere, J.J. and Swait, J. (2000). Attribute range effects in binary response tasks, *Marketing Letters* 11, 249–260.

Pullman, M.E., Dodson, K.J. and Moore, W.L. (1999). A comparison of conjoint methods when there are many attributes, *Marketing Letters* 10, 1–14.

Sándor, Z. and Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs, *Journal of Marketing Research* 38, 430–444.

Sándor, Z. and Wedel, M. (2002). Profile construction in experimental choice designs for mixed logit models, *Marketing Science* 21, 455–475.

Sándor, Z. and Wedel, M. (2005). Heterogeneous conjoint choice designs, *Journal of Marketing Research* 42, 210–218.

Scarpa, R., Campbell, D. and Hutchinson, W.G. (2007). Benefit transfer for landscape improvements: sequential bayesian design and respondent's rationality in a choice experiment, *Land Economics* 83, 617–634.

Scarpa, R., Ferrini, S. and Willis, K.G. (2005). Performance of error component models for status-quo effects in choice experiments, Ch13, in Scarpa, R. and Alberini, A. (eds), *Applications of Simulation Methods in Environmental and Resource Economics*, pp. 247–274. Springer, Publisher, Dordrecht, The Netherlands.

Scarpa, R., Thiene, M. and Train, K. (in press). Utility in WTP space: a tool to address

confounding random scale effects in destination choice to the Alps, *American Journal of Agricultural Economics*.

Street, D.J. and Burgess, L. (2005). Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments, *Journal of Statistical Planning and Inference* 118, 185–199.

Street, D.J., Bunch, D.S. and Moore, B.J. (2001). Optimal designs for 2^k paired comparison experiments, *Communications in Statistics, Theory, and Methods* 30, 2149–2171.

Street, D.J., Burgess, L. and Louviere, J.J. (2005). Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments, *International Journal of Research in Marketing* 22, 459–470.

Swait, J. and Louviere, J. (1993). The role of the scale parameter in the estimation and use of multinomiallogit models, *Journal of Marketing Research* 30, 305–314.

Toubia, O. and Hauser, J.R. (2007). On managerially efficient experimental designs, *Marketing Science* 25, 851–858.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, UK.

Train, K. and Weeks, M. (2005). Discrete choice models in preference space and willing-to-pay space, in Scarpa, R. and Alberini, A. (eds), *Applications of Simulation Methods in Environmental and Resource Economics*, Chapter 1, pp. 1–16. Springer Publisher, Dordrecht, The Netherlands.

Verlegh, P.W., Schifferstein, H.N. and Wittink, D.R. (2002). Range and Number-of-Levels Effects in Derived and Stated Measures of Attribute Importance, *Marketing Letters* 13, 41–52.