RESEARCH REPORT

COMPARING ALGORITHMS AND CRITERIA FOR DESIGNING BAYESIAN CONJOINT CHOICE EXPERIMENTS

ROSELINDE KESSELS • PETER GOOS • MARTINA VANDEBROEK

OR 0427

Comparing algorithms and criteria for designing Bayesian conjoint choice experiments

Roselinde Kessels
Peter Goos
Martina Vandebroek

Department of Applied Economics, Katholieke Universiteit Leuven, Belgium

Abstract

The recent algorithm to find efficient conjoint choice designs, the RSC-algorithm developed by Sándor and Wedel (2001), uses Bayesian design methods that integrate the $\mathcal{D}$-optimality criterion over a prior distribution of likely parameter values. Characteristic for this algorithm is that the designs satisfy the minimal level overlap property provided the starting design complies with it. Another, more embedded, algorithm in the literature, developed by Zwerina et al. (1996), involves an adaptation of the modified Fedorov exchange algorithm to the multinomial logit choice model. However, it does not take into account the uncertainty about the assumed parameter values. In this paper, we adjust the modified Fedorov choice algorithm in a Bayesian fashion and compare its designs to those produced by the RSC-algorithm. Additionally, we introduce a measure to investigate the utility balances of the designs. Besides the widely used $\mathcal{D}$-optimality criterion, we also implement the $\mathcal{A}$-, $\mathcal{G}$- and $\mathcal{V}$-optimality criteria and look for the criterion that is most suitable for prediction purposes and that offers the best quality in terms of computational effectiveness. The comparison study reveals that the Bayesian modified Fedorov choice algorithm provides more efficient designs than the RSC-algorithm and that the $\mathcal{D}$- and $\mathcal{V}$-optimality criteria are the best criteria for prediction, but the computation time with the $\mathcal{V}$-optimality criterion is longer.

Keywords: conjoint choice experiments, discrete choice experiments, Bayesian designs, multinomial logit, RSC-algorithm, Bayesian modified Fedorov choice algorithm, $\mathcal{D}$-, $\mathcal{A}$-, $\mathcal{G}$- and $\mathcal{V}$-optimality, predictive validity

1 Introduction

Since their breakthrough with the Louviere and Woodworth (1983) article, conjoint choice experiments, also referred to as discrete choice experiments, have become increasingly popular to explore consumer preferences for the attributes of various goods. In applied research, these experiments have been used extensively, while in fundamental research, they have been the subject of rigorous study and research. The reason for their popularity is that they enable researchers to model real marketplace choices, and thus to emulate real market decisions and predict market demand (Carson et al. 1994). In a typical conjoint choice experiment, respondents are presented with a series of choice sets, each composed of several alternatives, also called profiles, of products or services that are defined as combinations of different attribute levels. Respondents are then requested to indicate their preferred alternative for every choice set.

The aim of a conjoint choice experiment is to collect as much information as possible on the parameters of the corresponding statistical model by submitting a small number of choice sets to a limited number of respondents. To that end, an efficient experimental design needs to be developed. Specifically, designing a conjoint choice experiment consists of choosing the alternatives and grouping them into choice sets. In practice, the number of choice sets as well as the number of alternatives within a choice set and the set of all possible alternatives, i.e. the candidate alternatives, are first determined. The problem of finding an efficient choice design then comes down to selecting those alternatives that, when put into choice sets, provide the maximum of information on the model parameters. Assuming that the order of the choice sets in the design has no influence on the respondents' choices, the ranking of the choice sets in the design is unimportant.

In the literature on conjoint choice experiments, the efficiency of a design is commonly expressed in terms of the $\mathcal{D}$-optimality criterion (Atkinson and Donev 1992). This criterion represents the determinant of the information matrix of the parameter estimates. The design that maximizes this determinant, which is equivalent to minimizing the determinant of the variance-covariance matrix of the parameter estimates, is termed the $\mathcal{D}$-optimal design.

The main difficulty in the construction of a proper choice design is that the probabilistic choice models are nonlinear in the parameters, implying that the efficiency of the design depends on the unknown parameter vector (Atkinson and Haines 1996). Therefore, researchers need to assume values for the parameters before deriving their design. Naturally, if the true parameters are known, research is obviated.

The problem of how to design efficient choice experiments has been addressed several times in the literature. Primarily, the first steps were taken by Lazari and Anderson (1994) and Kuhfeld et al. (1994), who argue that $\mathcal{D}$-optimal designs for linear models also work well for the nonlinear choice models. In this way, they avoid the circular problem that the information on the parameters depends on the unknown values of those parame-

ters. Furthermore, they ignore the fact that choice design efficiency depends not only on the creation of appropriate alternatives, but also on properly placing them into various choice sets. Based on the $\mathcal{D}$-optimality criterion for the linear model, Lazari and Anderson (1994) provide a catalog of designs for logit choice models (McFaddden 1974). These designs perform well for certain problems. However, for many real problems, cataloged designs cannot be used without modification, and modification can reduce efficiency. To overcome this limitation, Kuhfeld et al. (1994) review a general computerized approach to compute designs for diverse situations. Specifically, they make use of Cook and Nachtsheim's (1980) modification of Fedorov's (1972) exchange algorithm to generate choice designs. Although this algorithm optimizes the $\mathcal{D}$-criterion for the linear model and not for the choice model, it has produced satisfactory choice designs for many years.

Thanks to Bunch et al. (1996), the choice design problem has been studied from a broader perspective. They start from the multinomial logit choice model (MNL, McFadden 1974) and assume zero prior parameter values. This means that all attribute levels, and thus all alternatives, are equally preferred by the respondents. Under this unrealistic assumption, the optimization problem for the MNL model again reduces to that for the linear model (Grossmann et al. 2002). The resulting choice designs are called $\mathcal{D}$-optimal utility-neutral or $\mathcal{D}_0$-optimal designs. Furthermore, Bunch et al. (1996) have developed a new, easy-to-use design strategy that yields a special class of $\mathcal{D}_0$-optimal designs for choice experiments. Their strategy is outlined in Section 4.1 of the paper. The designs they propose are the so-called shifted or cyclic designs characterized by the minimal level overlap property. This property is satisfied when the frequencies of the attribute levels within a choice set are distributed as equally as possible. It is regarded as an important property for choice designs since only the differences between attribute levels within a choice set are informative. Additionally, for main-effects MNL models, the shifted or cyclic designs are level balanced, i.e. the levels of an attribute occur with equal frequencies, and orthogonal, which implies that the number of times every two levels of an attribute occur together is the same for every pair of levels.

However, Huber and Zwerina (1996) have demonstrated that the assumption of zero prior values may be costly. In particular, if there are reasonable nonzero prior values, which is the case in most practical marketing research situations, then these can be used to generate choice designs with improved efficiency. The designs are then called $\mathcal{D}_P$-optimal designs, with subscript "P" referring to the nonzero prior point parameter values. In the literature on optimal design theory, these designs are termed locally optimal designs. The improvement in $\mathcal{D}_P$-efficiency occurs through two techniques, namely relabeling (R) and swapping (S). These techniques permute attribute levels and are described in detail in Section 4.1. Huber and Zwerina (1996) have found that generally, the utilities of the alternatives within the choice sets of $\mathcal{D}_P$-optimal main-effects designs are more balanced than the utilities in the $\mathcal{D}_0$-optimal shifted designs of Bunch et al. (1996) which they use as starting designs. The utilities are obtained by taking the sum of the weighted attribute levels for every alternative. According to the authors, choice sets with balanced utilities are to be preferred to choice sets with a dominating alternative. A dominating alter-

native in a choice set is an alternative of which the attributes are more attractive than those of the remaining, dominated, alternatives. In most cases, respondents immediately choose this dominating alternative and so the choice task is a very easy one. However, the choice of the respondents then does not provide much information on the parameters of the choice model and as a result, the parameters are estimated less precisely. To illustrate their point, Huber and Zwerina (1996) refer to the S-shaped cumulative probability function of the binary logit model. If one alternative within a pair is chosen almost all of the time, then this anchors the extreme values of the S-shaped cumulative probability function, but provides little information on the slope of that function. Moreover, they argue that the idea of utility balance is not new. Also the design heuristic for paired comparisons in the popular conjoint package, Adaptive Conjoint Analysis (ACA 1994), balances the alternatives to make respondents as indifferent to a pair as possible. To conclude, the authors have shown that the gains in $\mathcal{D}$-efficiency from bringing in a nonzero prior vector are relatively robust with respect to misspecifications of the prior values.

Another algorithm that enables researchers to integrate a likely prior parameter vector in the optimization process has been developed by Zwerina et al. (1996). They adapt the modified Fedorov exchange algorithm applied by Kuhfeld et al. (1994) to the specificities of the MNL model. We refer to the adapted version of the algorithm as the modified Fedorov choice algorithm. A more detailed description of the algorithm is given in Section 4.2. The main difference between the optimal designs generated by the modified Fedorov choice algorithm and by the RS-algorithm of Huber and Zwerina (1996) is that the former, contrary to the latter, are not restricted to level balance, orthogonality and minimal level overlap, although they approximately satisfy these constraints. Optimal designs that are generated without constraints are referred to as unconstrained optimal designs. Thus, the modified Fedorov choice algorithm yields unconstrained optimal designs.

The recent approach that produces efficient designs has been created by Sándor and Wedel (2001). As opposed to a.o. Huber and Zwerina (1996) and Zwerina et al. (1996), Sándor and Wedel (2001) only address main-effects models and thus no models with interactions. Their approach consists of integrating the associated uncertainty of the assumed parameter values by the use of Bayesian design techniques (Chaloner and Verdinelli 1995). If there is substantial uncertainty about the unknown parameters, then the resulting designs, the so-called Bayesian designs, outperform the designs proposed by Huber and Zwerina (1996) in terms of $\mathcal{D}$-efficiency. On the other hand, if the parameters are almost known for certain, which is seldom the case, then the designs proposed by Huber and Zwerina (1996) tend to do better. In particular, Sándor and Wedel (2001) determine a continuous prior distribution by eliciting prior information from respondents based on the methods developed by Van Lenthe (1993) and subsequently optimize the design over that distribution. Optimization occurs through the RS-algorithm, developed by Huber and Zwerina (1996), and an additional cycling (C) procedure. The resulting RSC-algorithm is discussed in Section 4.1. In contrast to the $\mathcal{D}_P$-optimal main-effects designs of Huber and Zwerina (1996), the produced designs, the so-called $\mathcal{D}_B$-optimal designs, are not characterized by perfect level balance. This is due to the cycling. Furthermore, Sándor

and Wedel (2001) state that orthogonality as design optimality constraint is irrelevant for nonlinear choice models since, in contrast to linear models, it does not imply that the parameter estimates themselves are independent. Therefore, they distinguish between design orthogonality, previously referred to as orthogonality, which applies to linear models, and information orthogonality, which applies to the nonlinear choice models. In case of information orthogonality, the parameters of the choice model can be estimated truly independently. Yet, for choice models, design efficiency depends on the unknown parameters and as a result, unless all alternatives are equally preferred, information orthogonality is unlikely to hold. Finally, Sándor and Wedel (2001) show that choice designs that yield more precise parameter estimates also improve predictive validity and thus the effectiveness of conjoint choice studies.

In the recent literature as summarized above, researchers have proposed a single (main-effects) design for the MNL model to be administered to different respondents whose choices are pooled. As a result, homogeneous parameters across respondents are assumed. Furthermore, the predictors, i.e. the choice attributes, are treated qualitative, or discrete. In this paper, we adopt the same experimental choice scenario and consider only main-effects choice designs.

The question whether the $\mathcal{D}_B$-optimal RSC-designs proposed by Sándor and Wedel (2001) reach unconstrained optimality or how close they come has not yet been answered in the literature. The modified Fedorov choice algorithm developed by Zwerina et al. (1996) produces unconstrained optimal designs, but has the disadvantage that it only integrates prior point estimates. In this paper, we extend this algorithm to incorporate the uncertainty about the prior estimates by using the same Bayesian procedure as in Sándor and Wedel (2001). In this way, the Bayesian optimal RSC and modified Fedorov choice designs can be easily compared, and as a result, the deviation from unconstrained optimality of the optimal RSC-designs can be determined. Additionally, we introduce a measure to examine the utility balances of the optimal designs. To evaluate design efficiency, we do not restrict ourselves to the widely used $\mathcal{D}$-optimality criterion, but we also integrate other optimality criteria in the algorithms. Specifically, the $\mathcal{A}$-, $\mathcal{G}$- and $\mathcal{V}$-optimality criteria, often used by researchers in optimal design, are addressed. We then investigate which of these design criteria leads to the most accurate estimates and predictions. Eventually, a conjoint choice experiment is carried out to predict the future market shares of the products or services under investigation. To summarize, Table 1 provides an overview of previous and current research on efficient choice designs.

The remainder of the paper is outlined as follows: the next section reviews the key ideas in generating optimal choice designs. Particularly, the multinomial logit formulation and the different optimality criteria are given. In Section 3, specific design measures are presented for evaluating the utility balance, the accuracy of the parameter estimates and the predictive validity. The RSC and Bayesian modified Fedorov choice algorithms are described in Section 4. In Section 5, the different algorithms and criteria are compared. Finally, Section 6 summarizes the paper and recalls the main conclusions.

Table 1: Search strategies for generating efficient aggregate choice designs when qualitative attributes and homogeneous parameters across respondents are assumed

Authors	Year	Crite-rion	Criterion definition	Assumed parameters	Algorithm
Lazari and Anderson	1994	$\mathcal{D}$	$\det(\mathbf{X}'\mathbf{X})$	—	not specified
Kuhfeld, Tobias and Garrat	1994	$\mathcal{D}$	$\det(\mathbf{X}'\mathbf{X})$	—	modified Fedorov
Huber and Zwerina	1996	$\mathcal{D}_P$	$\det\{\mathbf{I}(\mathbf{X},\boldsymbol{\beta})^{-1}\}$	$\boldsymbol{\beta}_0$	RS
Zwerina, Huber and Kuhfeld	1996	$\mathcal{D}_P$	$\det\{\mathbf{I}(\mathbf{X},\boldsymbol{\beta})^{-1}\}$	$\boldsymbol{\beta}_0$	modified Fedorov
Sándor and Wedel	2001	$\mathcal{D}_B$	$\det\{\mathbf{I}(\mathbf{X},\boldsymbol{\beta})^{-1}\}$	$N(\boldsymbol{\beta}\|\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$	RSC
This paper	2004	$\mathcal{D}_B$ $\mathcal{A}_B$ $\mathcal{G}_B$ $\mathcal{V}_B$	$\det\{\mathbf{I}(\mathbf{X},\boldsymbol{\beta})^{-1}\}$ $\mathrm{tr}\{\mathbf{I}(\mathbf{X},\boldsymbol{\beta})^{-1}\}$ $\max_{\mathbf{x_{jd}}} \mathbf{c}'(\mathbf{x_{jd}})\mathbf{I}(\mathbf{X},\boldsymbol{\beta})^{-1}\mathbf{c}(\mathbf{x_{jd}})$ $\int \mathbf{c}'(\mathbf{x_{jd}})\mathbf{I}(\mathbf{X},\boldsymbol{\beta})^{-1}\mathbf{c}(\mathbf{x_{jd}})\mathrm{d}\mathbf{x_{jd}}$	$\boldsymbol{\beta} \sim U[\text{-}1, 1]^k$	RSC modified Fedorov

Note: R = relabeling, S = swapping, C = cycling; for an explanation, see Section 4.1.

2 Theoretical framework

The designs we develop in this paper are based on the main-effects MNL model derived from McFadden's (1974) random utility model of consumer choice. In the random utility model, a person's utility for an alternative j in a choice set is specified as

$$u_j = \mathbf{x_j}'\boldsymbol{\beta} + \varepsilon_j, \tag{1}$$

where $\mathbf{x_j}$ is the $k \times 1$ vector of the attributes of alternative j, $\boldsymbol{\beta}$ the $k \times 1$ vector of parameters $[\beta_1, \ldots, \beta_k]'$ weighting these attributes, and ε_j the i.i.d. standard Gumbel, or type 1 extreme value (1, 0), error term.

Suppose there are S choice sets, C_s, indexed by $s = 1, 2, ..., S$, where each choice set is characterized by a set of J alternatives: $C_s = \{\mathbf{x_{1s}}, ..., \mathbf{x_{Js}}\}$. If N respondents have to choose the alternative that maximizes their perceived utility of the J alternatives in choice set s, then the multinomial logit probability that alternative j is chosen can be expressed in closed form as

$$p_{js}(\mathbf{X_s}, \boldsymbol{\beta}) = \frac{e^{\mathbf{x_{js}}'\boldsymbol{\beta}}}{\sum_{t=1}^{J} e^{\mathbf{x_{ts}}'\boldsymbol{\beta}}}, \tag{2}$$

where $\mathbf{X_s} = [\mathbf{x_{1s}}, ..., \mathbf{x_{Js}}]'$. The stacked $\mathbf{X_s}$ matrices provide the design matrix $\mathbf{X}$ for the choice experiment.

Because of the assumption of independent error terms, the choices from N respondents made among the alternatives in the S choice sets represent independent draws from a multinomial distribution. Therefore, if $\mathbf{Y} = [\mathbf{y_1}, ..., \mathbf{y_N}]$ denotes the matrix of choices from N respondents with elements y_{jsn}, each of which equals one if respondent n, $n = 1, 2, ..., N$, chooses alternative j in C_s and zero otherwise, then the log-likelihood of the N samples in $\mathbf{Y}$ is defined as

$$\ln\{\mathrm{L}(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta})\} = \sum_{s=1}^{S} \sum_{j=1}^{J} \sum_{n=1}^{N} y_{jsn} \ln\{p_{js}(\mathbf{X_s}, \boldsymbol{\beta})\}. \tag{3}$$

Maximizing this expression with respect to $\boldsymbol{\beta}$ yields the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ for a particular choice design.

The derivation of optimal choice designs is based on the Fisher information matrix, which is inversely proportional to the variance-covariance matrix of the parameter estimates, and given by

$$\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}) = N \sum_{s=1}^{S} \mathbf{X_s'}(\mathbf{P_s} - \mathbf{p_s}\mathbf{p_s'})\mathbf{X_s}, \tag{4}$$

where $\mathbf{p_s} = [p_{1s}, ..., p_{Js}]'$ and $\mathbf{P_s} = \mathrm{diag}[p_{1s}, ..., p_{Js}]$. In Appendix A, we show how the information matrix is obtained from the log-likelihood function in (3). The composition of a choice design $\mathbf{X}$ clearly influences the information content or the precision or quality of the choice experiment. Also, since $\mathbf{p_s}$ and $\mathbf{P_s}$ both depend on $\boldsymbol{\beta}$ according to the multinomial logit model in (2), the circular problem that the information on the parameters depends on their unknown values can be observed. By maximizing the information on the unknown parameters, optimal choice designs are generated. To that end, optimal design theory (Fedorov 1972, Silvey 1980, Atkinson and Donev 1992) proposes a set of optimality criteria that minimize a function of the information matrix in (4). Additionally, each of these criteria addresses a slightly different requirement for the design. In the sequel of this section, we present an overview of the optimality criteria most often encountered in the optimal design literature. These criteria belong to the class of alphabetic optimality criteria because they are called after a letter. We first give a brief definition of each criterion in general and then apply it to the multinomial logit model in (2) to derive locally and Bayesian optimality criteria. However, it is important to point out that only Bayesian optimal designs are computed in this paper (see Section 4) as they usually outperform locally optimal designs.

$\mathcal{D}$-optimality

As described in the introduction, researchers in conjoint choice experiments have only used the $\mathcal{D}$-optimality criterion to develop locally and Bayesian optimal choice designs. This widely accepted one-dimensional measure of information seeks to maximize the determinant of the information matrix in (4), or to minimize its inverse, the determinant of the variance-covariance matrix of the parameter estimates, which is often called the

generalized variance. For local optimality, researchers in conjoint choice experiments usually refer to the $\mathcal{D}$-optimality criterion as the $\mathcal{D}_P$-optimality criterion, to make the distinction with linear models for which no prior point estimate is required. Formally, the $\mathcal{D}_P$-criterion value is given by

$$\mathcal{D}_P = \{\text{det}\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1}\}^{1/k}, \tag{5}$$

where the exponent $1/k$ assures that it is independent of the dimension k of the parameter vector $\boldsymbol{\beta}$. The $\mathcal{D}_P$-optimal design is thus obtained by minimizing the $\mathcal{D}_P$-criterion value given a prior parameter vector $\boldsymbol{\beta_0}$.

To adopt a Bayesian design approach, one needs to specify the prior distribution $\pi(\boldsymbol{\beta})$ of the logit coefficients. This distribution can be informative or uninformative depending on the availability of prior information. The Bayesian optimal design is then obtained by minimizing the $\mathcal{D}_B$-optimality criterion value, i.e. the expectation of the $\mathcal{D}_P$-optimality criterion value over the prior distribution $\pi(\boldsymbol{\beta})$. Formally, the $\mathcal{D}_B$-criterion value is defined as

$$\mathcal{D}_B = \text{E}_{\boldsymbol{\beta}}\left\{\{\text{det}\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1}\}^{1/k}\right\} = \int_{\mathcal{R}^k} \{\text{det}\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1}\}^{1/k} \pi(\boldsymbol{\beta}) \text{d}\boldsymbol{\beta}. \tag{6}$$

In practice, the $\mathcal{D}_B$-criterion value is approximated by drawing R prior parameter values $\boldsymbol{\beta}^r$, $r = 1, ..., R$, from $\pi(\boldsymbol{\beta})$, and computing

$$\widetilde{\mathcal{D}}_B(\mathbf{X}) = \frac{1}{R} \sum_{r=1}^{R} \{\text{det}\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^r)^{-1}\}^{1/k}. \tag{7}$$

The motivation for $\mathcal{D}$-optimality originates from the confidence ellipsoid for $\boldsymbol{\beta}$ which has a volume inversely proportional to the square root of the determinant of the information matrix $\mathbf{I}$. Hence, the $\mathcal{D}_P$- and $\mathcal{D}_B$-optimal designs minimize the (expected) volume of the ellipsoid and thus minimize both the variances and covariances, or the generalized variance, of the parameter estimates.

In the literature on conjoint choice experiments, the $\mathcal{D}$-optimality criterion has been advocated for computing optimal choice designs because it has some advantages over other alphabetic optimality criteria. The underpinnings of these advantages mainly stem from optimal design theory for linear models. First, the $\mathcal{D}_P$- and $\mathcal{D}_B$-optimal designs usually perform well with regard to other optimality criteria, whereas this is often not true for other optimality criteria. Second, the $\mathcal{D}_P$- and $\mathcal{D}_B$-optimality criteria are invariant to a linear transformation of the design matrix. As a result, there are invariant to the scale or coding of the attributes. This means that the rank order of the efficiency of a number of designs remains unchanged when different codings of the attributes, e.g. dummy or effects coding, are used. Finally, the $\mathcal{D}_P$-criterion value is computationally efficient to update in exchange algorithms thanks to the existence of a simple, but powerful formula for

recalculating the determinant of the information matrix $\mathbf{I}$. The use of this update formula is illustrated in Zwerina et al. (1996). However, the update formula is not applicable for constructing the $\mathcal{D}_B$-optimal design due to the presence of the prior parameter distribution $\pi(\boldsymbol{\beta})$.

$\mathcal{A}$-optimality

The $\mathcal{A}$-optimality criterion prefers the design for which the sum or the average of the variances of the parameter estimates is minimized. For the nonlinear choice model in (2), the $\mathcal{A}_P$-criterion value equals

$$\mathcal{A}_P = \sum_{i=1}^{k} \text{var}(\hat{\beta}_i) = \text{tr}\{\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1}\}. \tag{8}$$

The $\mathcal{A}_P$-optimal design refers to the design for which the $\mathcal{A}_P$-criterion value is minimized given a reasonable prior parameter vector $\boldsymbol{\beta_0}$. The Bayesian version of the $\mathcal{A}_P$-criterion value, termed the $\mathcal{A}_B$-criterion value, is obtained by averaging the $\mathcal{A}_P$-criterion value over the prior distribution $\pi(\boldsymbol{\beta})$ of the coefficients. Hence, the $\mathcal{A}_B$-optimal design minimizes

$$\mathcal{A}_B = \text{E}_{\boldsymbol{\beta}}\left\{\text{tr}\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1}\right\} = \int_{\mathcal{R}^k} \text{tr}\{\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1}\}\pi(\boldsymbol{\beta})\text{d}\boldsymbol{\beta}, \tag{9}$$

or by approximation,

$$\widetilde{\mathcal{A}}_B(\mathbf{X}) = \frac{1}{R}\sum_{r=1}^{R} \text{tr}\{\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^r)^{-1}\}, \tag{10}$$

with R the number of draws from $\pi(\boldsymbol{\beta})$.

Notice that the $\mathcal{A}_P$- and $\mathcal{A}_B$-optimality criteria only consider the parameter variances as opposed to the $\mathcal{D}_P$- and $\mathcal{D}_B$-optimality criteria which take both the variances and covariances into account.

An advantage of the $\mathcal{A}$-optimality criterion with regard to other design criteria is that the coefficients can be weighted. However, a drawback of this criterion is that it is not invariant to recodings of the attributes, i.e. the ordering of designs with respect to the $\mathcal{A}_P$- or $\mathcal{A}_B$-optimality criteria depends on the type of coding. We refer to Goos (2002) for an example in the case of linear models.

$\mathcal{G}$-optimality

The $\mathcal{G}$-optimal design minimizes the maximal prediction variance over the design region χ. For the nonlinear choice model in (2), the design region χ consists of all possible choice sets of size J that can be composed from the candidate alternatives $\mathbf{x_j}$. If there are D

possible choice sets, $C_d = \{\mathbf{x_{1d}}, ..., \mathbf{x_{Jd}}\}$, indexed by $d = 1, 2, ..., D$, then the $\mathcal{G}_P$-criterion value equals

$$\mathcal{G}_P = \max_{\mathbf{x_{jd}} \in \chi} \text{var}\{\hat{p}_{jd}(\mathbf{x_{jd}}, \boldsymbol{\beta})\} = \max_{\mathbf{x_{jd}} \in \chi} \mathbf{c}'(\mathbf{x_{jd}}) \mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1} \mathbf{c}(\mathbf{x_{jd}}), \tag{11}$$

where $\hat{p}_{jd}(\mathbf{x_{jd}}, \boldsymbol{\beta})$ denotes the predicted choice probability for $\mathbf{x_{jd}}$ and

$$\mathbf{c}(\mathbf{x_{jd}}) = \frac{\partial p_{jd}(\mathbf{x_{jd}}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \tag{12}$$

the first-order truncated Taylor series expansion of the multinomial logit model in (2). This approach is similar to the computation of locally $\mathcal{D}$- and c-optimal designs for non-linear models in general (Atkinson and Donev 1992, Atkinson and Haines 1996). Based on model (2), (12) can be written as

$$\begin{aligned}
\frac{\partial p_{jd}(\mathbf{x_{jd}}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{e^{\mathbf{x}'_{\mathbf{jd}}\boldsymbol{\beta}} \mathbf{x_{jd}} \sum_{t=1}^{J} e^{\mathbf{x}'_{\mathbf{td}}\boldsymbol{\beta}} - e^{\mathbf{x}'_{\mathbf{jd}}\boldsymbol{\beta}} \sum_{t=1}^{J} e^{\mathbf{x}'_{\mathbf{td}}\boldsymbol{\beta}} \mathbf{x_{td}}}{\left(\sum_{t=1}^{J} e^{\mathbf{x}'_{\mathbf{td}}\boldsymbol{\beta}}\right)^2} \\
&= \frac{e^{\mathbf{x}'_{\mathbf{jd}}\boldsymbol{\beta}}}{\sum_{t=1}^{J} e^{\mathbf{x}'_{\mathbf{td}}\boldsymbol{\beta}}} \left(\frac{\mathbf{x_{jd}} \sum_{t=1}^{J} e^{\mathbf{x}'_{\mathbf{td}}\boldsymbol{\beta}} - \sum_{t=1}^{J} e^{\mathbf{x}'_{\mathbf{td}}\boldsymbol{\beta}} \mathbf{x_{td}}}{\sum_{t=1}^{J} e^{\mathbf{x}'_{\mathbf{td}}\boldsymbol{\beta}}} \right) \\
&= p_{jd} \left\{ \mathbf{x_{jd}} - \sum_{t=1}^{J} \left(\frac{e^{\mathbf{x}'_{\mathbf{td}}\boldsymbol{\beta}}}{\sum_{v=1}^{J} e^{\mathbf{x}'_{\mathbf{vd}}\boldsymbol{\beta}}} \right) \mathbf{x_{td}} \right\} \\
&= p_{jd} \left(\mathbf{x_{jd}} - \sum_{t=1}^{J} p_{td} \mathbf{x_{td}} \right).
\end{aligned} \tag{13}$$

The design that minimizes the $\mathcal{G}_P$-criterion value given a prior parameter vector $\boldsymbol{\beta_0}$ is referred to as the $\mathcal{G}_P$-optimal design.

The $\mathcal{G}_B$-optimal design is obtained by minimizing the $\mathcal{G}_B$-criterion value, i.e. the average $\mathcal{G}_P$-criterion value over the prior parameter distribution $\pi(\boldsymbol{\beta})$. More specifically, the $\mathcal{G}_B$-optimal design minimizes

$$\mathcal{G}_B = \text{E}_{\boldsymbol{\beta}} \left\{ \max_{\mathbf{x_{jd}} \in \chi} \mathbf{c}'(\mathbf{x_{jd}}) \mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1} \mathbf{c}(\mathbf{x_{jd}}) \right\} = \int_{\mathcal{R}^k} \max_{\mathbf{x_{jd}} \in \chi} \mathbf{c}'(\mathbf{x_{jd}}) \mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1} \mathbf{c}(\mathbf{x_{jd}}) \pi(\boldsymbol{\beta}) \text{d}\boldsymbol{\beta}, \tag{14}$$

which is approximated by

$$\widetilde{\mathcal{G}}_B(\mathbf{X}) = \frac{1}{R} \sum_{r=1}^{R} \max_{\mathbf{x_{jd}} \in \chi} \mathbf{c}'(\mathbf{x_{jd}}) \mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^r)^{-1} \mathbf{c}(\mathbf{x_{jd}}). \tag{15}$$

For not too small design problems, the computation of the $\mathcal{G}_B$-optimal design by minimizing the $\mathcal{G}_B$-criterion value as defined in (14) is demanding. This is because the specification of the design region χ as the collection of all possible choice sets of size J is cumbersome. As a result, the computation of the $\mathcal{G}_B$-optimal design is lengthy and time-consuming. Therefore, we will use a tighter and more tractable specification of the design region and refer to it as $\chi*$. In particular, we will specify the design region $\chi*$ as one large choice set containing all candidate alternatives and modify the $\mathcal{G}_B$-optimality criterion accordingly.

$\mathcal{V}$-optimality

The $\mathcal{V}$-optimal design minimizes the average prediction variance over the design region χ. This criterion is sometimes referred to as $\mathcal{Q}$-optimality, as $\mathcal{I}$-optimality or as $\mathcal{IV}$-optimality as well. Like the $\mathcal{G}_P$-criterion value, the $\mathcal{V}_P$-criterion value is formally defined on the design region χ that consists of all D possible choice sets of size J that can be composed from the candidate alternatives $\mathbf{x_j}$. The $\mathcal{V}_P$-criterion value then equals

$$\mathcal{V}_P = \int_{\chi} \mathbf{c}'(\mathbf{x_{jd}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1}\mathbf{c}(\mathbf{x_{jd}})\mathrm{d}\mathbf{x_{jd}}, \tag{16}$$

with $\mathbf{c}(\mathbf{x_{jd}})$ defined as in (12) and (13).

The locally optimal design which minimizes the $\mathcal{V}_P$-criterion value is referred to as the $\mathcal{V}_P$-optimal design. Averaging the $\mathcal{V}_P$-criterion value over the prior parameter distribution $\pi(\boldsymbol{\beta})$ yields the $\mathcal{V}_B$-criterion value which is given by

$$\begin{aligned} \mathcal{V}_B &= \mathrm{E}_{\boldsymbol{\beta}} \left\{ \int_{\chi} \mathbf{c}'(\mathbf{x_{jd}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1}\mathbf{c}(\mathbf{x_{jd}})\mathrm{d}\mathbf{x_{jd}} \right\} \\ &= \int_{\mathcal{R}^k} \int_{\chi} \mathbf{c}'(\mathbf{x_{jd}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})^{-1}\mathbf{c}(\mathbf{x_{jd}})\mathrm{d}\mathbf{x_{jd}}\pi(\boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta}, \end{aligned} \tag{17}$$

and approximated by

$$\widetilde{\mathcal{V}}_B(\mathbf{X}) = \frac{1}{R}\frac{1}{D}\frac{1}{J}\sum_{r=1}^{R}\sum_{d=1}^{D}\sum_{j=1}^{J} \mathbf{c}'(\mathbf{x_{jd}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^r)^{-1}\mathbf{c}(\mathbf{x_{jd}}). \tag{18}$$

The $\mathcal{V}_B$-optimal design is obtained by minimizing the $\mathcal{V}_B$-criterion value given the prior distribution $\pi(\boldsymbol{\beta})$.

Similar to the computation of the $\mathcal{G}_B$-optimal design, we will base the computation of the $\mathcal{V}_B$-optimal design on the more workable design region $\chi*$ consisting of one large choice set that comprises the candidate alternatives.

In Appendix B, we present a simple numerical example of constructing Bayesian optimal designs by means of the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimality criteria.

3 Specific design measures

In this section, we discuss some specific, practical measures on which we evaluate the Bayesian optimal designs. We first introduce a measure for the utility balance of a design and then describe some measures for the accuracy of the parameter estimates and the predictive validity.

3.1 Utility balance

The measure we introduce to compare the utility balances of the Bayesian optimal choice designs is the percentage utility balance. To compute the percentage utility balance of a Bayesian optimal design, we first multiply the so-called within-choice set probabilities, i.e. the probabilities of the alternatives within a choice set derived from the multinomial logit model in (2) for a prior parameter from $\pi(\boldsymbol{\beta})$. We repeat the multiplications for several prior parameters from $\pi(\boldsymbol{\beta})$ and average them out. Formally, for the S choice sets C_s, the outcome M_s of averaging multiplications of the probabilities for choice set s can be expressed as

$$M_s = \int_{\mathcal{R}^k} \left(\prod_{j=1}^{J} p_{js}(\mathbf{X_s}, \boldsymbol{\beta}) \right) \pi(\boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta}, \tag{19}$$

or by approximation,

$$\widetilde{M}_s = \frac{1}{R} \sum_{r=1}^{R} \left(\prod_{j=1}^{J} p_{js}(\mathbf{X_s}, \boldsymbol{\beta}^r) \right). \tag{20}$$

We then add up the values for $\widetilde{M}_s$ over the S choice sets to obtain a final value B given by

$$B = \sum_{s=1}^{S} \widetilde{M}_s. \tag{21}$$

For a perfectly utility balanced design, all within-choice set probabilities are equal to $1/J$ so that $M_s = (1/J)^J$ for every choice set resulting in a maximum value for B, B_{max}, given by

$$B_{max} = S \left(\frac{1}{J} \right)^J. \tag{22}$$

Finally, B_{max} can be used to express the value for B of a design with an imperfect utility balance as a percentage. The percentage utility balance $\%B$ then becomes

$$\%B = \frac{B}{B_{max}} \times 100, \tag{23}$$

and indicates the extent to which a design is utility balanced.

3.2 Accuracy of the parameter estimates

In this section, we propose two one-dimensional measures to compare the Bayesian optimal designs with regard to their estimation performance of the parameters. The first measure is the expected mean square error of the parameter estimates ($\mathrm{EMSE}_{\hat{\boldsymbol{\beta}}}$) and the second measure is the so-called average standard error of the parameter estimates ($\bar{s}_{\hat{\boldsymbol{\beta}}}$). The $\mathcal{A}_B$-optimal design is expected to score best on the latter measure as it minimizes the expected

sum or average of the parameter variances, and thus does not take the covariances into account.

EMSE of the parameter estimates

Formally, the EMSE of the parameter estimates is given by

$$\text{EMSE}_{\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta}) = \int_{\mathcal{R}^k} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) f(\hat{\boldsymbol{\beta}}) \mathrm{d}\hat{\boldsymbol{\beta}}, \tag{24}$$

where $f(\hat{\boldsymbol{\beta}})$ is the distribution of the estimates. It can be shown that the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$ captures both the variances of the estimates and the biases. The smaller the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$-value, the more precisely the parameters are estimated.

In practice, the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$ is approximated by computing R times $\hat{\boldsymbol{\beta}}^r$, $r = 1, ..., R$, through simulation and working out

$$\widetilde{\text{EMSE}}_{\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta}) = \frac{1}{R} \sum_{r=1}^{R} \left(\hat{\boldsymbol{\beta}}^r - \boldsymbol{\beta}\right)' \left(\hat{\boldsymbol{\beta}}^r - \boldsymbol{\beta}\right). \tag{25}$$

Starting from a true parameter vector $\boldsymbol{\beta}$, the simulation process proceeds as follows. To obtain an estimate $\hat{\boldsymbol{\beta}}$, we simulate choices y_{jsn} for N respondents with respect to a Bayesian optimal design by drawing for each choice set s of the design and for each respondent n a random number v_{sn} from the uniform distribution $U[0, 1]$. These random numbers represent cumulative probabilities and are compared to the true logit probabilities $\mathbf{p_s}(\boldsymbol{\beta})$ of the design to assign values to y_{jsn} in the following way:

$$y_{jsn} = \begin{cases} 1 & \text{if } v_{sn} \in \left]\sum_{t=1}^{j-1} p_{ts}(\boldsymbol{\beta}), \sum_{t=1}^{j} p_{ts}(\boldsymbol{\beta})\right], \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

Then, based on the vectors $\mathbf{y_n}$, we estimate $\hat{\boldsymbol{\beta}}$ using the log-likelihood function in (3).

Average standard error of the parameter estimates

Another measure to assess the quality of the parameter estimates $\hat{\boldsymbol{\beta}}^r$ obtained from a Bayesian optimal design is the average standard error of the parameter estimates, $\bar{s}_{\hat{\boldsymbol{\beta}}}$. This measure is derived from the variances of the k parameter estimates, $s^2_{\hat{\beta}_i}$, $i = 1, ..., k$, which are defined as

$$s^2_{\hat{\beta}_i}(\boldsymbol{\beta}) = \frac{1}{R-1} \sum_{r=1}^{R} \left(\hat{\beta}_i^r - \overline{\hat{\beta}}_i\right)^2, \tag{27}$$

where R refers to the number of simulations and $\overline{\hat{\beta}}_i$ is the average parameter estimate over the R simulations. The average standard error of the parameter estimates, $\bar{s}_{\hat{\boldsymbol{\beta}}}$, is then given by

$$\bar{s}_{\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta}) = \sqrt{\frac{1}{k} \sum_{i=1}^{k} s^2_{\hat{\beta}_i}}. \tag{28}$$

3.3 Predictive validity

In general, marketing consultants carry out conjoint choice experiments to predict the market share for certain products or services as precisely as possible. The market shares are derived from the predicted probabilities that are calculated using the multinomial logit formulation in (2). To assist marketing consultants in making precise predictions, we evaluate the predictive validity of the Bayesian optimal designs by means of two measures that are analogous to those used for evaluating the accuracy of the parameter estimates. The first measure is the expected mean square error of the predicted probabilities ($\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$) and the second measure is the average standard error of the predicted probabilities ($\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$). We apply both measures to a special design for prediction, namely a design that consists of all possible choice sets of size J. In this way, no particular design is favoured. We refer to this special design as the complete choice design.

In the optimal design literature, the $\mathcal{G}$- and $\mathcal{V}$-optimality criteria are proposed to generate designs for making precise predictions about the response variable. It is therefore interesting to investigate the relative predictive validity of the popular $\mathcal{D}_B$-optimal design with respect to the $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal designs. These issues are further elaborated in Section 5 of the paper.

EMSE of the predicted probabilities

Formally, the EMSE of the predicted probabilities is given by

$$\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}(\boldsymbol{\beta}) = \int_{\mathcal{R}^k} \left(\hat{\mathbf{p}}_\mathbf{c}(\hat{\boldsymbol{\beta}}) - \mathbf{p}_\mathbf{c}(\boldsymbol{\beta})\right)' \left(\hat{\mathbf{p}}_\mathbf{c}(\hat{\boldsymbol{\beta}}) - \mathbf{p}_\mathbf{c}(\boldsymbol{\beta})\right) f(\hat{\boldsymbol{\beta}}) \mathrm{d}\hat{\boldsymbol{\beta}}. \tag{29}$$

Here, $\mathbf{p}_\mathbf{c}(\boldsymbol{\beta})$ is the vector of logit probabilities in the complete choice design based on the true parameter $\boldsymbol{\beta}$, $\hat{\mathbf{p}}_\mathbf{c}(\hat{\boldsymbol{\beta}})$ is the corresponding vector of predicted logit probabilities computed for the parameter estimate $\hat{\boldsymbol{\beta}}$ and $f(\hat{\boldsymbol{\beta}})$ is the distribution of the estimates. Consequently, the $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$ compares the true and predicted probabilities using the true and estimated parameters, respectively. The smaller the $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$-value, the more precisely choices are predicted.

To use the $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$ in practice, it is approximated by

$$\widetilde{\text{EMSE}}_{\hat{\mathbf{p}}_\mathbf{c}}(\boldsymbol{\beta}) = \frac{1}{R}\sum_{r=1}^{R} \left(\hat{\mathbf{p}}_\mathbf{c}(\hat{\boldsymbol{\beta}}^r) - \mathbf{p}_\mathbf{c}(\boldsymbol{\beta})\right)' \left(\hat{\mathbf{p}}_\mathbf{c}(\hat{\boldsymbol{\beta}}^r) - \mathbf{p}_\mathbf{c}(\boldsymbol{\beta})\right), \tag{30}$$

where R refers to the number of simulations.

Average standard error of the predicted probabilities

Another way to look at the predictive validity of a Bayesian optimal design is to study the average standard error of the predicted probabilities in the complete choice design, $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$. If there are D choice sets in the complete choice design, indexed by $d = 1, 2, ..., D$, then,

to compute this measure, we first compute the variances of the predicted probabilities of the alternatives in the D choice sets, $s^2_{\hat{p}_{c_{jd}}}$, which are given by

$$s^2_{\hat{p}_{c_{jd}}}(\boldsymbol{\beta}) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{p}_{c_{jd}}(\hat{\boldsymbol{\beta}}^r) - \overline{\hat{p}}_{c_{jd}}\right)^2, \tag{31}$$

where $\overline{\hat{p}}_{c_{jd}}$ is the average predicted probability of alternative j in choice set d over the R simulations. The computation of $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$ then comes down to

$$\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}(\boldsymbol{\beta}) = \sqrt{\frac{1}{D}\frac{1}{J}\sum_{d=1}^{D}\sum_{j=1}^{J} s^2_{\hat{p}_{c_{jd}}}}. \tag{32}$$

4 Bayesian design generating algorithms

This section embarks on the discussion of the algorithms that are used for constructing $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal choice designs. The algorithms under investigation are the RSC-algorithm and the Bayesian modified Fedorov choice algorithm. In both algorithms, we implement the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimality criterion because Bayesian optimal designs are usually more efficient than locally optimal designs. The design we chose to optimize with the different algorithms and optimality criteria is characterized by four attributes each at three levels, and two alternatives in each of fifteen choice sets. The design is therefore said to be of type $3^4/2/15$. We now start with a description of the RSC-algorithm and then proceed with the account of the Bayesian modified Fedorov choice algorithm.

4.1 The RSC-algorithm

The RSC-algorithm comprises three sub-algorithms, namely relabeling (R), swapping (S) and cycling (C), hereby constructing a so-called RSC-design. Relabeling and swapping have been introduced by Huber and Zwerina (1996) in combination with the use of nonzero prior point coefficients. Cycling has been added by Sándor and Wedel (2001). Additionally, these authors have adapted the RSC-algorithm in a Bayesian manner. In this paper, we deal with this recent version of the RSC-algorithm.

Starting design

As starting design for the RSC-algorithm, we took the same starting design as in Sándor and Wedel (2001, Table 2, p. 434–435), which is depicted in Table 2. Noteworthy about this starting design is that it is characterized by minimal level overlap. In the literature, minimal level overlap is considered as a property of efficient choice designs and if the starting design is characterized by it, then the designs generated by the algorithm

Table 2: Starting design

Choice set	Profile	Attributes 1	2	3	4
1	I	3	2	3	2
	II	1	3	2	1
2	I	3	1	2	1
	II	1	2	1	3
3	I	3	3	2	3
	II	1	1	1	2
4	I	3	2	1	1
	II	1	3	3	3
5	I	3	1	1	2
	II	1	2	3	1
6	I	1	1	2	2
	II	2	2	1	1
7	I	1	2	3	3
	II	2	3	2	2
8	I	1	1	1	1
	II	2	2	3	3
9	I	1	3	1	3
	II	2	1	3	2
10	I	1	3	3	1
	II	2	1	2	3
11	I	2	2	2	2
	II	3	3	1	1
12	I	2	3	3	1
	II	3	1	2	3
13	I	2	2	2	3
	II	3	3	1	2
14	I	2	3	1	2
	II	3	1	3	1
15	I	2	1	3	3
	II	3	2	2	2

also satisfy this property. Hence, the algorithm limits its search to designs with the minimal level overlap property provided the starting design complies with it. Also the starting designs proposed by Huber and Zwerina (1996) satisfy the minimal level overlap constraint. These designs are the $\mathcal{D}_0$-optimal shifted or cyclic designs developed by Bunch et al. (1994). They are constructed by first generating a so-called seed design by use of either an orthogonal array or the SAS procedure OPTEX. Each row of this design represents the first alternative of a particular choice set. The other alternatives of the choice sets are generated by subsequently incrementing the levels of the attributes with each new alternative and cycling back to the lowest level when the highest level for an attribute is reached. In this way, the designs attain the minimal level overlap property.

In general, to construct and evaluate designs with algorithms, the levels of the attributes are coded. We made use of the so-called effects-type coding which is often employed in the literature. In our situation of three levels per attribute, the first level of each attribute is coded as [1 0], the second level as [0 1] and the third level as [-1 -1]. The use of this type of coding explains why only two parameter values are needed to characterize each of the four attributes. Hence, in our situation, the number of parameter values, k, equals 8.

Prior parameter distribution

The prior distribution from which we sampled to compute the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-criterion value is the continuous multivariate uniform distribution, $\pi(\boldsymbol{\beta}) = U[\text{-}1, 1]^k$. This distribution gives equal weight to all possible prior parameter vectors $\boldsymbol{\beta_0}$ in the interval $[\text{-}1, 1]^k$. Consequently, it is useful when there is substantial uncertainty about the unknown parameters. We fixed the bounds on -1 and 1 because we experienced that the use of higher absolute values leads to designs that consist of many choice sets with a dominating alternative. Particularly when the Bayesian modified Fedorov choice algorithm was used, which does not confine its search to designs with the minimal level overlap property, it could be seen that large absolute prior parameter values entailed optimal designs with a lot of overlap in the choice sets. Most often, overlap was found in three of the four attributes so that dominating alternatives were clearly present in the designs.

Similar to Sándor and Wedel (2001), we draw $R = 1000$ prior parameter samples $\boldsymbol{\beta}^r$ to compute the Bayesian criterion values. A relatively large number of draws is needed, since for a small number of draws, chance fluctuations affect the computation of the criterion values and thus also the construction of the optimal designs.

Relabeling

Relabeling modifies a design by permuting the levels of the attributes across choice sets and searches for a combination of permutations for which the corresponding design has the highest efficiency. This design is called the optimal R-design. Take for example the starting design of the algorithm in Table 2. If ceteris paribus the levels 1, 2 and 3 of attribute 1 are replaced by 3, 1 and 2, and the levels 1, 2 and 3 of attribute 4 by 2, 1 and 3,

then the corresponding design is retained if it improves efficiency. Consequently, relabeling the starting design involves an investigation of $3! \times 3! \times 3! \times 3! = 1296$ permutations because each of the four attributes has three levels for which 3! permutations are possible. The optimal R-design then refers to the design with the smallest criterion value among the 1296 designs.

Swapping

Swapping involves switching two attribute levels among alternatives within a choice set and testing if the swap improves the criterion value. If this is the case, the swap is performed. Furthermore, the procedure also considers simultaneous swaps for several attributes within a choice set. In the end, if no design is found with a lower criterion value, then the last selected design is the one that maximizes the information and is called the optimal RS-design. Starting from the optimal R-design of our example with two alternatives, the swapping algorithm begins with the first choice set, takes the level of the first attribute and swaps it with the level of that attribute in the second alternative. Then it does the same with the other attributes after which simultaneous swaps are examined. The algorithm subsequently proceeds to the second choice set and passes through all choice sets. If an improvement in information occurs, then the modified design is retained and the algorithm returns to the first choice set and continues in the same way until improvement stops. In that case, the optimal RS-design is obtained.

Cycling

Cycling is a combination of cyclically rotating the levels of an attribute and swapping them. All cycles and swaps are examined for design improvements and if an improvement emerges, the corresponding design is stored. If, after a while, no improvement is possible, then the last stored design is the one with the largest efficiency gains and is called the optimal RSC-design. Applied to the optimal RS-design of our example, the cycling procedure starts with the first attribute in the first choice set. It cyclically rotates the levels of the attribute until all possibilities are exhausted. Thus, for the three attribute levels, level 1 is replaced by level 2, level 2 by level 3 and level 3 by level 1. This goes on until the original configuration is obtained again. Then a swap is applied and the attribute levels are cyclically rotated for the second time. For designs with more than two alternatives, this alternating swapping and cycling continues until all swaps and subsequent cycles are verified. Then the algorithm proceeds to the second choice set and so on until the last choice set. Subsequently, the remaining attributes are handled in the same way as the first attribute. At each stage, if an improvement is made, the algorithm starts over from the first attribute in the first choice set. Finally, the algorithm stops when no further improvement is achieved and the last selected design is the optimal RSC-design.

4.2 The Bayesian modified Fedorov choice algorithm

Contrary to the RSC-algorithm, the Bayesian modified Fedorov choice algorithm is flexible in the sense that its design search is not restricted to minimal level overlap designs.

The designs generated are referred to as the Bayesian optimal modified Fedorov choice designs. We used the same multivariate uniform prior parameter distribution as the one employed in the RSC-algorithm, namely $\pi(\boldsymbol{\beta}) = U[\text{-}1, 1]^k$, and we also took the same $R = 1000$ draws to compute the criterion values.

The Bayesian modified Fedorov choice algorithm is an exchange algorithm. This means that the design alternatives are exchanged with the alternatives from a predefined set of candidate profiles. The algorithm originally stems from the Fedorov algorithm (1972) developed for constructing optimal designs for linear models. The Fedorov algorithm carries out an exchange if it is the best exchange possible. To speed up the algorithm, Cook and Nachtsheim (1980) modified it by executing any beneficial exchange as soon as it is discovered rather than only performing the best exchange. Finally, Zwerina et al. (1996) adapted the modified Fedorov algorithm to the multinomial logit model. While they only incorporated one prior point parameter vector in the optimization process, we adapted the modified Fedorov algorithm in a Bayesian fashion. We now explain the algorithm more in detail and proceed with a study of the effectiveness of the different design criteria used for the algorithm.

Description of the algorithm

The Bayesian modified Fedorov choice algorithm starts with the composition of the set of candidate profiles. For our example, this set consists of $3^4 = 81$ candidates. Next, the starting choice design is constructed by randomly selecting alternatives from this set. The algorithm alters the starting design by exchanging its alternatives with the candidates. Specifically, for every alternative in the starting design, the exchange with every candidate alternative is considered and if the exchange improves the efficiency of the corresponding design with respect to a certain criterion, then it is performed. Remark that only exchanges are considered for which the candidate alternative is different from the alternatives in the choice set tackled. The first iteration is completed when the algorithm has found the best exchanges for the alternatives in the starting design. After that, the algorithm returns to the first alternative and continues with iterations until no substantial efficiency improvement is possible anymore. In total, an iteration in our example requires testing $30 \times (81 - 2) = 2370$ exchanges. To avoid poor local optima, the algorithm repeats the search for a number of different starting designs. Each repetition is called a try and we executed 150 tries. Finally, the most efficient design from these tries is selected and is called the Bayesian optimal modified Fedorov choice design.

Computational effectiveness of the design criteria

The computational effectiveness of a Bayesian design criterion refers to the quality and the speed of the Bayesian modified Fedorov choice algorithm in which this criterion is integrated. We compared the computational effectiveness of the Bayesian design criteria by means of the estimated expected efficiency of the corresponding optimal designs. The computation of expected efficiencies was first done by Trinca and Gilmour (2000) in the context of blocked experiments. If M refers to the number of tries for the algorithm,

then, to compute the expected efficiency of a Bayesian optimal modified Fedorov design, the efficiencies E_m of the designs $\mathbf{X}^m$, $m = 1, ..., M$, that are generated by try m of the algorithm first need to be derived. As already mentioned above, we used $M = 150$ tries. The efficiencies are given by

$$E_m = \frac{\mathcal{B}(\mathbf{X}^*)}{\mathcal{B}(\mathbf{X}^m)}, \tag{33}$$

where $\mathcal{B}$ refers to the corresponding Bayesian criterion value, $\mathcal{D}_B$, $\mathcal{A}_B$, $\mathcal{G}_B$ or $\mathcal{V}_B$, and $\mathbf{X}^*$ to the respective optimal design.

Let's assume now that for a large number of tries, t, we obtain g distinct designs G_1, $G_2, ..., G_g$, with efficiencies $E_1 > E_2 > ... > E_g$. G_1 is the best design, and an estimate of the probability of finding G_1 in t tries, say ρ_1, is given by the number of times G_1 is found divided by t. If, correspondingly, ρ_2, ..., ρ_g refer to the probability of finding $G_2, ..., G_g$ in t tries, then the estimated expected efficiency of the Bayesian optimal design is defined as

$$\text{E(efficiency)} = \sum_{i=1}^{g-1} \left\{ \left(\sum_{j=i}^{g} \rho_j \right)^t - \left(\sum_{j=i+1}^{g} \rho_j \right)^t \right\} E_i + \rho_g^t E_g. \tag{34}$$

Usually, this expression is computed for various numbers of tries, t, to make a plot of the expected efficiencies. In total, we made four plots, one for each of the Bayesian optimality criteria. From these plots, we will determine the optimality criterion for which the algorithm obtains the optimum with the least number of tries. This criterion also leads to the highest expected efficiency if the optimum is missed.

5 Results

In this section, we compare the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal designs generated by the RSC and Bayesian modified Fedorov choice algorithms. First, the RSC-algorithm is compared to the Bayesian modified Fedorov choice algorithm for each criterion. Specifically, the designs, their computation time, utility balance and performance on local optimality are studied. Next, the different optimality criteria are compared in terms of other design criteria than the one optimized for, the accuracy of the parameter estimates, the predictive validity and the computational effectiveness. We performed all computations in SAS 8.02, procedure IML.

5.1 Comparing the RSC and Bayesian modified Fedorov choice algorithms

The RSC and Bayesian modified Fedorov choice algorithms are now compared with respect to the optimal designs generated, the corresponding criterion values, the computation times, the degree of utility balance and their performance on local optimality.

Optimal designs and criterion values

The Bayesian optimal RSC and modified Fedorov designs are depicted in Tables 3 and 4 respectively. The optimal RSC-designs clearly do not show any overlap because of the minimal level overlap constraint integrated in the RSC-algorithm if the starting design complies with it, which is the case here. Since there is no restriction to minimal level overlap in the Bayesian modified Fedorov choice algorithm, some overlap can be seen in the $\mathcal{G}_B$-optimal modified Fedorov design and one overlap in the $\mathcal{V}_B$-optimal modified Fedorov design.

Table 5 shows the results of the comparison between the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal RSC and modified Fedorov designs. A close look at this table reveals that for every design criterion used, the Bayesian optimal modified Fedorov designs are more efficient than the optimal RSC-designs. This is because the Bayesian modified Fedorov choice algorithm is not restricted in its design search as opposed to the RSC-algorithm. From Table 5 we also notice that with regard to the RSC-algorithm the swapping of attribute levels within a choice set has no impact on lowering the criterion values of the optimal R-designs. On the other hand, the relabeling and cycling of attribute levels are effective in achieving better designs.

Thus, one better constructs optimal choice designs by means of the Bayesian modified Fedorov choice algorithm because this algorithm generates more efficient designs. However, if one uses the RSC-algorithm to optimize criterion values, then Table 6 indicates by how much the optimal criterion values deviate from the unconstrained optimal criterion values, i.e. resulting from the Bayesian modified Fedorov choice algorithm. The values of the $\mathcal{D}_B$- and $\mathcal{V}_B$-optimal RSC-designs only deviate by 1.72% and 1.67% from the unconstrained optimal values, whereas those of the $\mathcal{A}_B$- and $\mathcal{G}_B$-optimal RSC-designs show a deviation of 4.72% and 6.28% respectively. As a result, the $\mathcal{D}_B$- and $\mathcal{V}_B$-optimal designs generated by the RSC-algorithm come close to unconstrained optimality, but the $\mathcal{A}_B$- and $\mathcal{G}_B$-optimal designs are somewhat further away from it.

Computation time

In Table 7, computation times are specified for the Bayesian optimal RSC and modified Fedorov choice designs. The designs were generated in SAS 8.02, PROC IML. The times were obtained using a Dell PC with a 1.80 GHz Intel Processor and 256 MB RAM. For the Bayesian optimal modified Fedorov designs, times are reported for one try, i.e. using one random starting design, so as to compare with the optimal RSC-designs for which the starting design of Table 2 was used. Furthermore, the times for computing the modified Fedorov choice designs are approximate since the number of iterations in one try is not fixed, but determined by convergence. From the table, we clearly observe that the computation times for the Bayesian optimal modified Fedorov designs are much higher than those for the optimal RSC-designs.

Table 3: Bayesian optimal RSC-designs

		$\mathcal{D}_B$-optimal design				$\mathcal{A}_B$-optimal design				$\mathcal{G}_B$-optimal design				$\mathcal{V}_B$-optimal design			
Choice set	Profile	Attributes				Attributes				Attributes				Attributes			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	I	2	2	2	1	2	1	2	2	1	2	2	1	1	1	2	3
	II	1	1	3	2	1	2	3	1	2	1	3	2	2	2	3	2
2	I	2	1	2	2	2	2	2	1	1	2	2	2	1	2	3	2
	II	1	2	1	1	1	1	1	2	2	1	1	1	2	1	1	1
3	I	2	1	3	1	2	2	2	2	1	2	3	1	1	2	3	1
	II	1	2	1	3	1	1	1	1	2	1	1	3	2	1	1	3
4	I	2	2	1	2	2	1	1	1	1	1	1	2	1	1	1	2
	II	1	1	2	1	1	2	2	2	2	2	2	1	2	2	2	1
5	I	2	1	1	1	2	2	1	2	1	3	1	3	1	2	1	1
	II	1	2	2	2	1	1	2	1	2	2	2	2	2	1	2	2
6	I	1	3	2	3	1	3	2	3	2	3	2	3	2	3	2	3
	II	3	2	1	2	3	1	1	1	3	2	1	2	3	1	1	2
7	I	1	3	1	1	1	1	1	2	1	1	1	1	2	3	1	1
	II	3	1	3	3	3	2	3	3	2	2	3	3	3	1	3	3
8	I	1	3	1	2	1	3	1	1	2	3	1	2	2	3	1	2
	II	2	2	2	1	2	1	2	2	1	1	2	1	1	1	2	1
9	I	1	2	1	1	1	2	1	2	2	2	1	1	2	1	1	1
	II	3	3	2	3	3	3	2	3	3	3	2	3	3	3	2	3
10	I	1	1	2	2	1	2	2	1	2	1	2	2	2	2	2	2
	II	3	3	1	3	3	3	1	2	3	3	1	1	3	3	1	1
11	I	1	2	3	3	1	1	2	3	2	1	3	3	2	1	3	3
	II	2	1	1	2	2	2	1	1	1	2	1	2	1	2	2	2
12	I	3	1	2	2	3	2	2	1	3	1	2	2	3	2	2	2
	II	2	3	3	1	2	3	3	2	2	3	3	1	1	3	3	1
13	I	3	2	3	1	3	1	3	2	3	1	3	1	3	1	3	1
	II	2	1	1	3	2	2	1	3	2	2	1	3	1	2	1	3
14	I	3	1	1	3	3	2	1	3	3	2	1	3	3	2	1	3
	II	2	3	2	2	2	3	2	1	2	3	2	2	1	3	2	2
15	I	3	3	2	1	3	3	2	2	3	3	2	2	3	3	2	1
	II	2	2	3	3	2	1	3	3	1	2	3	3	1	1	3	3
		$\widetilde{\mathcal{D}}_B =$ 0.32479				$\widetilde{\mathcal{A}}_B =$ 3.25170				$\widetilde{\mathcal{G}}_B =$ 0.01340				$\widetilde{\mathcal{V}}_B =$ 0.00058			

Table 4: Bayesian optimal modified Fedorov designs

		$\mathcal{D}_B$-optimal design				$\mathcal{A}_B$-optimal design				$\mathcal{G}_B$-optimal design				$\mathcal{V}_B$-optimal design			
Choice set	Profile	Attributes				Attributes				Attributes				Attributes			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	I	2	2	2	3	1	1	3	2	3	3	1	2	3	1	3	2
	II	1	1	3	2	2	2	1	1	2	1	3	3	3	2	1	1
2	I	3	1	3	2	1	2	2	1	2	3	2	3	2	2	3	1
	II	1	3	2	1	2	3	1	2	3	1	1	2	1	1	1	2
3	I	3	1	2	2	2	1	2	2	2	3	1	2	2	1	1	1
	II	2	2	1	1	1	2	3	1	1	1	2	1	3	2	2	2
4	I	1	3	1	3	2	2	1	3	1	2	1	2	1	2	2	1
	II	3	2	2	1	3	1	2	1	2	3	1	1	2	1	1	3
5	I	1	3	2	2	2	1	1	1	2	2	2	2	1	1	3	1
	II	2	1	1	3	1	3	2	2	1	1	3	1	2	2	2	2
6	I	2	1	2	1	2	2	2	1	1	3	1	2	2	2	3	2
	II	1	2	1	2	3	1	1	2	3	3	3	1	3	3	1	3
7	I	2	2	3	1	3	2	1	2	3	2	1	3	3	3	1	1
	II	3	1	1	3	1	3	3	3	2	1	3	2	1	2	3	3
8	I	3	3	1	2	3	2	2	1	3	2	2	1	3	3	3	3
	II	1	1	3	3	1	1	1	2	3	1	1	3	2	1	2	2
9	I	1	2	2	3	2	3	3	1	2	2	1	3	2	2	1	2
	II	2	1	1	1	1	1	2	3	1	3	2	3	1	3	2	1
10	I	1	2	2	1	2	2	3	2	1	2	3	3	3	1	2	2
	II	2	3	1	3	1	3	2	1	2	1	2	1	1	3	1	3
11	I	1	1	1	1	3	2	3	3	3	1	2	2	3	3	3	2
	II	2	3	3	2	2	1	1	1	1	2	2	1	1	2	2	3
12	I	3	2	1	1	1	2	1	3	2	2	3	2	2	3	1	2
	II	2	1	2	2	2	1	3	1	1	1	1	1	3	1	2	3
13	I	2	2	1	2	3	3	1	3	1	1	2	2	1	1	1	1
	II	3	3	2	1	1	2	2	2	3	2	1	1	2	3	2	3
14	I	2	2	2	2	2	2	2	2	2	1	1	1	1	3	2	2
	II	1	3	3	1	1	1	1	1	3	1	2	3	3	2	1	1
15	I	1	1	1	2	2	1	2	3	2	1	2	3	1	2	1	2
	II	3	2	3	3	1	2	1	2	1	3	3	2	2	1	2	1
		$\widetilde{\mathcal{D}}_B =$ 0.31930				$\widetilde{\mathcal{A}}_B =$ 3.10527				$\widetilde{\mathcal{G}}_B =$ 0.01261				$\widetilde{\mathcal{V}}_B =$ 0.00057			

Table 5: $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-criterion values of the RSC starting design and the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal designs generated by relabeling, swapping and cycling (RSC) and the Bayesian modified Fedorov choice algorithm

Optimal design	RSC				Modified Fedorov
	Start*	Relabeling	Swapping	Cycling	
$\mathcal{D}_B$	0.35533	0.34776	0.34776	0.32479	0.31930
$\mathcal{A}_B$	4.22441	4.05209	4.05209	3.25170	3.10527
$\mathcal{G}_B$	0.01711	0.01548	0.01548	0.01340	0.01261
$\mathcal{V}_B$	0.00072	0.00067	0.00067	0.00058	0.00057

*Start refers to the starting design in Table 2

Table 6: Deviation from unconstrained optimality of the Bayesian optimal RSC-designs

Optimal design	% Deviation of RSC from modified Fedorov
$\mathcal{D}_B$	1.72%
$\mathcal{A}_B$	4.72%
$\mathcal{G}_B$	6.28%
$\mathcal{V}_B$	1.67%

Utility balance

Table 8 indicates the percentage utility balances of the RSC starting design and the optimal designs generated by relabeling, swapping and cycling (RSC) and the Bayesian modified Fedorov choice algorithm. In general, the percentage values of the optimal choice designs are very low and thus, the designs are not utility balanced. With regard to the RSC-algorithm, it can be seen that the relabeling and cycling increase the percentage utility balance to some extent. Further, there is no substantial difference in utility balance between the Bayesian optimal RSC and modified Fedorov designs.

Local optimality

It is interesting to investigate the performance of the Bayesian optimal RSC and modified Fedorov choice designs for particular values of $\boldsymbol{\beta}$. Specifically, we assume that a certain parameter is the true parameter $\boldsymbol{\beta}$. For two true parameters, we computed local criterion values for the Bayesian optimal designs and investigated to what extent the local criterion values of the designs deviate from the true optimal criterion values. Therefore, we computed the locally $\mathcal{D}_P$-, $\mathcal{A}_P$-, $\mathcal{G}_P$- and $\mathcal{V}_P$-optimal RSC and modified Fedorov designs. The first parameter used was randomly drawn from $U[-1, 1]^k$ whereas the second parameter contains all zero values and was chosen as a limiting case. The results for the two parameters are shown in Table 9. A first, general finding is that the local criterion values

Table 7: Computation times of the Bayesian optimal RSC and modified Fedorov designs. For the modified Fedorov designs, approximate computations times are given for one try.

Optimal design	RSC	Modified Fedorov
$\mathcal{D}_B$	0.08h	1h
$\mathcal{A}_B$	0.08h	1h
$\mathcal{G}_B$	1.06h	7h
$\mathcal{V}_B$	1.06h	7h

Table 8: Percentage utility balances of the RSC starting design and the optimal designs generated by relabeling, swapping and cycling (RSC) and the Bayesian modified Fedorov choice algorithm

Optimal design	RSC				Modified Fedorov
	Start*	Relabeling	Swapping	Cycling	
$\mathcal{D}_B$	55.30%	55.93%	55.93%	57.69%	57.83%
$\mathcal{A}_B$	55.30%	55.52%	55.52%	59.21%	59.18%
$\mathcal{G}_B$	55.30%	55.44%	55.44%	58.00%	59.57%
$\mathcal{V}_B$	55.30%	55.44%	55.44%	57.12%	57.96%

*Start refers to the starting design in Table 2

for the Bayesian optimal modified Fedorov designs are smaller or equal than those for the optimal RSC-designs. As a result, the Bayesian modified Fedorov choice algorithm produces better designs than the RSC-algorithm for these parameters. Further, it is remarkable that the deviations with respect to the $\mathcal{G}_P$-optimality criterion are much higher than those with respect to the other criteria.

5.2 Comparing the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimality criteria

Since for every design criterion used, the Bayesian optimal modified Fedorov designs are more efficient than the optimal RSC-designs, we compared the different design criteria only with respect to the Bayesian modified Fedorov choice algorithm. In this section, we first elaborate on their performance in terms of other optimality criteria. Next, we consider the accuracy of the parameter estimates generated by the optimal designs and the predictive validity of the designs. Finally, we discuss the computational effectiveness of the design criteria.

Table 9: Local $\mathcal{D}_P$-, $\mathcal{A}_P$-, $\mathcal{G}_P$- and $\mathcal{V}_P$-criterion values of the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal designs generated by the RSC and Bayesian modified Fedorov choice algorithms. For each of the local criterion values, the percentage deviation from the corresponding optimal local criterion value is indicated.

a) Values are given for the true parameter
$\boldsymbol{\beta} = [0.25847, -0.54128, 0.15842, -0.48284, -0.18535, 0.64285, -0.12391, 0.33847]'$

Optimal design	RSC		Modified Fedorov	
	Local criterion value	% Deviation from local optimum	Local criterion value	% Deviation from local optimum
$\mathcal{D}_B$	0.23606	24.13%	0.22376	19.32%
$\mathcal{A}_B$	2.18020	17.85%	2.07095	14.50%
$\mathcal{G}_B$	0.00284	141.60%	0.00230	129.59%
$\mathcal{V}_B$	0.00023	17.13%	0.00023	22.46%

b) Values are given for the true parameter $\boldsymbol{\beta} = [0, 0, 0, 0, 0, 0, 0, 0]'$

Optimal design	RSC		Modified Fedorov	
	Local criterion value	% Deviation from local optimum	Local criterion value	% Deviation from local optimum
$\mathcal{D}_B$	0.16142	3.17%	0.16004	3.31%
$\mathcal{A}_B$	1.40934	1.31%	1.39018	3.24%
$\mathcal{G}_B$	0.00020	18.81%	0.00020	55.97%
$\mathcal{V}_B$	0.00012	3.42%	0.00012	8.06%

Performance in terms of other optimality criteria

Since the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimality criteria all have a slightly different aim, it is interesting to observe how robust the Bayesian optimal modified Fedorov designs are with respect to other design criteria for which they are not optimized. Table 10 gives the efficiencies of the Bayesian optimal designs with respect to the different optimality criteria. As expected from optimal design theory, the $\mathcal{D}_B$-optimal design scores quite well on the other design criteria. In particular, the $\mathcal{D}_B$-optimal design is almost as good as the $\mathcal{V}_B$-optimal design in terms of $\mathcal{V}_B$-efficiency. Hence, the $\mathcal{D}_B$-optimal design is not only useful for precisely estimating the parameters, but also for accurately predicting logit probabilities. Inversely, the $\mathcal{V}_B$-optimal design also performs well in terms of $\mathcal{D}_B$-efficiency. As a result, the $\mathcal{D}_B$- and $\mathcal{V}_B$-optimal designs behave similarly. To a lesser extent, this is also the case for the $\mathcal{D}_B$- and $\mathcal{A}_B$-optimal designs since the corresponding optimality criteria are related functions of the information matrix in (4). Further, it is striking that the $\mathcal{G}_B$-optimal design does not perform well on the other design criteria. Also the $\mathcal{D}_B$- and $\mathcal{A}_B$-optimal designs do not score well in terms of $\mathcal{G}_B$-efficiency, in contrast with the $\mathcal{V}_B$-optimal design.

Table 10: Performance of the Bayesian optimal modified Fedorov designs in terms of other design criteria

Design criterion	Optimal design			
	$\mathcal{D}_B$	$\mathcal{A}_B$	$\mathcal{G}_B$	$\mathcal{V}_B$
$\mathcal{D}_B$	100.00%	98.26%	91.67%	98.50%
$\mathcal{A}_B$	98.13%	100.00%	93.06%	96.66%
$\mathcal{G}_B$	96.24%	93.97%	100.00%	99.20%
$\mathcal{V}_B$	99.73%	96.27%	96.78%	100.00%

Accuracy of the parameter estimates

To evaluate the accuracy of the parameter estimates generated by the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal modified Fedorov designs, we computed for each of these designs the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$ and the average standard error of the parameter estimates, $\bar{s}_{\hat{\boldsymbol{\beta}}}$. Since both measures depend on a true parameter $\boldsymbol{\beta}$, we repeated the computation of these measures 50 times, each time for a different true parameter. We refer to each new computation for a different true parameter as a replication. For each replication, we estimated the parameter values $R = 1000$ times by simulating choices from $N = 20$ respondents.

Table 11 shows the results of the 50 replications. By means of percentage values, it depicts the number of replications for which the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal modified Fedorov designs have the lowest value for the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$ and the average standard error of the parameter estimates, $\bar{s}_{\hat{\boldsymbol{\beta}}}$. Overall, for each of the 50 replications, the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$ and the average standard error of the parameter estimates, $\bar{s}_{\hat{\boldsymbol{\beta}}}$, lead to similar conclusions about the ranking of the optimal designs with respect to the best estimate. As expected, the $\mathcal{A}_B$-optimal design has the largest number of replications for which the estimate $\hat{\boldsymbol{\beta}}$ is most precise. Furthermore, we observe that the estimation performance of the $\mathcal{D}_B$-optimal design is fairly good and that of the $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal designs somewhat poorer. As an illustration, we display the outcomes of a replication in Table 12a. For this replication, the $\mathcal{A}_B$-optimal design finds the best estimate of the true parameter $\boldsymbol{\beta}$.

It is also interesting to depict the values of the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$ and the average standard error of the parameter estimates, $\bar{s}_{\hat{\boldsymbol{\beta}}}$, for the 50 replications by means of box plots in Figures 1a and 1b. The white line in each of the boxes is the median. Both figures, which respectively display the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$-distributions and the $\bar{s}_{\hat{\boldsymbol{\beta}}}$-distributions, show similar patterns. Surprisingly, we spot the lowest median for the $\mathcal{D}_B$-optimal design, but the medians for the other optimal designs, in particular the $\mathcal{A}_B$-optimal design, are close. Furthermore, the interquartile range in all four box plots is comparable. This is also the case for the distance between the whiskers, although it is a bit smaller for the $\mathcal{G}_B$-optimal design. Generally, we can conclude from this section that, although the $\mathcal{A}_B$-optimal design has the largest number of replications for which the estimate $\hat{\boldsymbol{\beta}}$ is the best, the

Table 11: Number of replications for which the Bayesian optimal modified Fedorov designs perform best on providing the most accurate estimate $\hat{\boldsymbol{\beta}}$ in terms of the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$ and the average standard error of the parameter estimates, $\bar{s}_{\hat{\boldsymbol{\beta}}}$

Optimal design	# Replications with lowest $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$	$\bar{s}_{\hat{\boldsymbol{\beta}}}$
$\mathcal{D}_B$	30%	32%
$\mathcal{A}_B$	42%	44%
$\mathcal{G}_B$	14%	12%
$\mathcal{V}_B$	14%	12%

Table 12: a) Values for the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$ and the average standard error of the parameter estimates, $\bar{s}_{\hat{\boldsymbol{\beta}}}$, indicating the estimation performance of the parameters of the Bayesian optimal modified Fedorov designs. b) Values for the $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$ and the average standard error of the predicted probabilities, $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$, indicating the predictive validity of the Bayesian optimal modified Fedorov designs. Values are given for the true parameter $\boldsymbol{\beta} = [0.45654, 0.65320, 0.18420, -0.34280, -0.98530, -0.10420, 0.64280, -0.12390]'$.

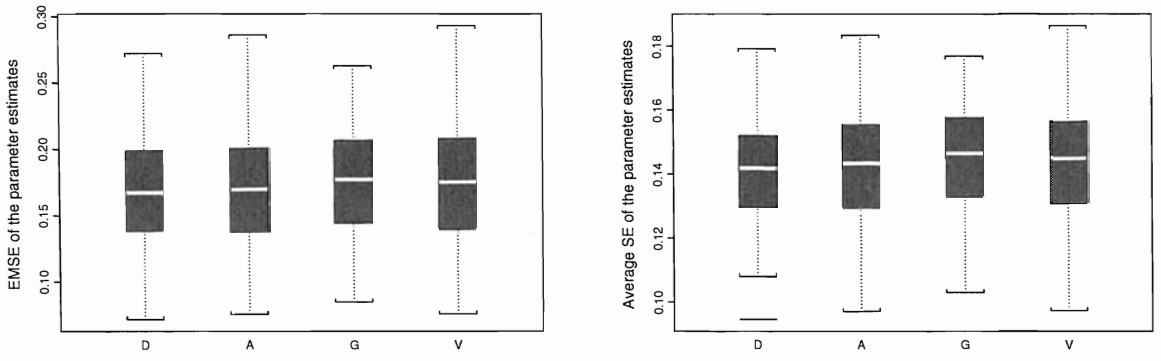
		Optimal design			
		$\mathcal{D}_B$	$\mathcal{A}_B$	$\mathcal{G}_B$	$\mathcal{V}_B$
a)	$\text{EMSE}_{\hat{\boldsymbol{\beta}}}$	0.18293	0.16749	0.21164	0.18807
	$\bar{s}_{\hat{\boldsymbol{\beta}}}$	0.14873	0.14196	0.16033	0.15131
b)	$\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$	29.72293	31.31291	38.01960	30.07124
	$\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$	0.06765	0.06942	0.07657	0.06808

$\mathcal{D}_B$-optimality criterion comes very close to the $\mathcal{A}_B$-optimality criterion for providing accurate estimates.

Predictive validity

In this section, we finally establish which of the Bayesian optimal modified Fedorov designs provides the best predictions $\hat{\mathbf{p}}_\mathbf{c}(\hat{\boldsymbol{\beta}})$ with respect to the complete choice design. To this end, we computed for each of the optimal designs the $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$ and the average standard error of the predicted probabilities, $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$. We carried out 50 replications and for each replication, we computed the predicted probabilities $R = 1000$ times using $N = 20$ respondents to simulate choices.

The prediction results from the 50 replications are shown in Table 13. The table depicts the number of replications for which the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal modified Fedorov designs have the lowest value for the $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$ and the average standard error of the predicted probabilities, $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$. Surprisingly, not the $\mathcal{V}_B$-optimal design, but the $\mathcal{D}_B$-



(a) $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$-distributions

(b) $\bar{s}_{\hat{\boldsymbol{\beta}}}$-distributions

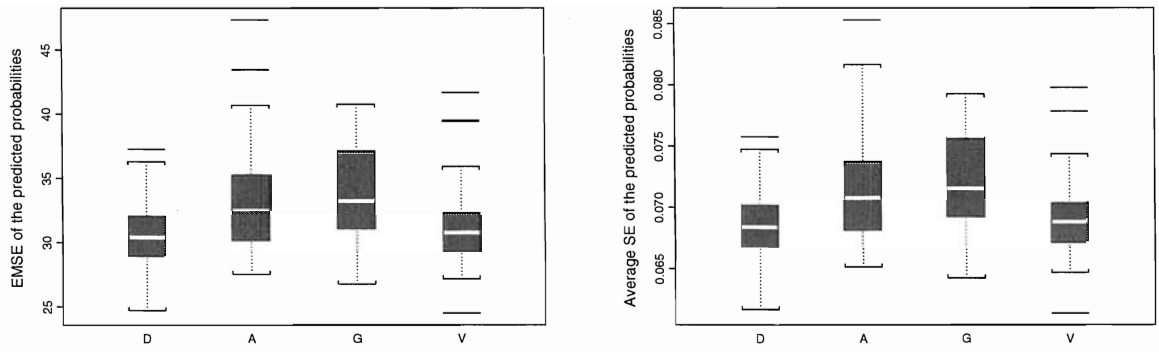
Figure 1: Distributions of the $\text{EMSE}_{\hat{\boldsymbol{\beta}}}$ and the average standard error of the parameter estimates, $\bar{s}_{\hat{\boldsymbol{\beta}}}$, obtained from 50 replications and computed for the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal modified Fedorov designs

optimal design has the largest number of replications for which the predictions $\hat{\mathbf{p}}_\mathbf{c}(\hat{\boldsymbol{\beta}})$ are most precise. Nevertheless, the $\mathcal{D}_B$-optimal design is closely followed by the $\mathcal{V}_B$-optimal design. The $\mathcal{A}_B$- and $\mathcal{G}_B$-optimal designs provide the best predictions in only a few replications. As an example, the outcomes of a replication are portrayed in Table 12b. For this replication, the $\mathcal{D}_B$-optimal design provides the most accurate predictions.

Table 13: Number of replications for which the Bayesian optimal modified Fedorov designs perform best on making the most precise predictions $\hat{\mathbf{p}}_\mathbf{c}(\hat{\boldsymbol{\beta}})$ with respect to the complete choice design in terms of the $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$ and the average standard error of the predicted probabilities, $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$

Optimal design	# Replications with lowest $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$ and $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$
$\mathcal{D}_B$	46%
$\mathcal{A}_B$	12%
$\mathcal{G}_B$	8%
$\mathcal{V}_B$	34%

In Figures 2a and 2b, the distributions of the values for the $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$ and the average standard error of the predicted probabilities, $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$, for the 50 replications are depicted by means of box plots. It is striking that the medians and the interquartile ranges of the boxes for the $\mathcal{D}_B$- and $\mathcal{V}_B$-optimal designs are the same in both figures. The distance between the whiskers of the box plots for the $\mathcal{V}_B$-optimal design is smaller than that for the $\mathcal{D}_B$-optimal design, although more outliers are present for the $\mathcal{V}_B$-optimal design. Concerning



(a) $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$-distributions

(b) $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$-distributions

Figure 2: Distributions of the $\text{EMSE}_{\hat{\mathbf{p}}_\mathbf{c}}$ and the average standard error of the predicted probabilities, $\bar{s}_{\hat{\mathbf{p}}_\mathbf{c}}$, obtained from 50 replications and computed for the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal modified Fedorov designs

the distributions for the $\mathcal{A}_B$- and $\mathcal{G}_B$-optimal designs, we observe rather elevated values for the median, a large interquartile range and a large distance between the whiskers. To conclude, we can state that the $\mathcal{D}_B$-optimal design most often leads to the best predictions. However, the predictions with respect to the $\mathcal{V}_B$-optimal design are almost as precise as those with respect to the $\mathcal{D}_B$-optimal design.

Computational effectiveness of the design criteria

To compare the computational effectiveness of the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimality criterion when applied in the Bayesian modified Fedorov choice algorithm, we computed the estimated expected efficiencies of the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal modified Fedorov designs according to (34). The results are shown in Figure 3. It turns out that the highest expected efficiency is obtained when the $\mathcal{D}_B$-optimality criterion is used. The $\mathcal{V}_B$-optimality criterion is second-best, followed by the $\mathcal{G}_B$-optimality criterion and the $\mathcal{A}_B$-optimality criterion. As a consequence, the smallest number of tries is needed for calculating the $\mathcal{D}_B$-optimal modified Fedorov design. Moreover, missing the optimal $\mathcal{D}_B$-criterion value leads to the smallest loss in efficiency. To summarize, choosing the $\mathcal{D}_B$-optimality criterion in combination with the Bayesian modified Fedorov choice algorithm to obtain an efficient choice design not only yields very accurate predictions, but also results in the highest efficiency if the optimal design is missed.

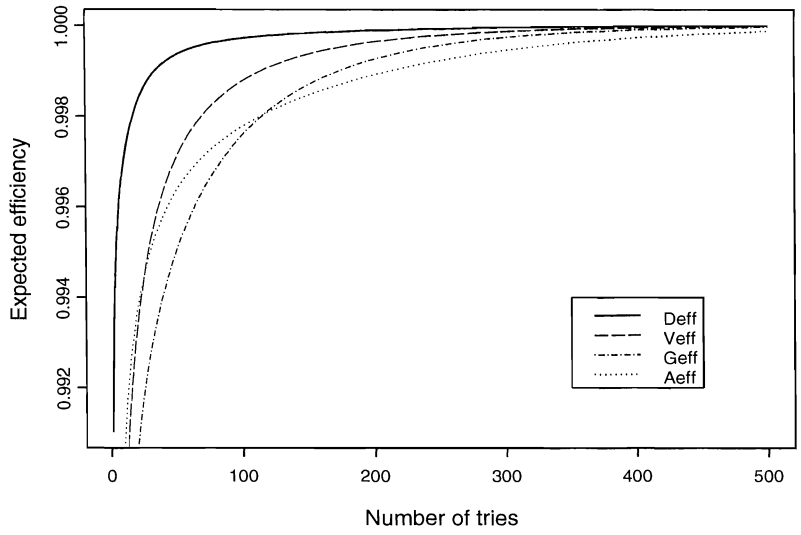


Figure 3: Estimated expected efficiency of the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimal modified Fedorov designs

6 Summary and conclusion

In this paper, we adapted the modified Fedorov choice algorithm developed by Zwerina et al. (1996) in a Bayesian manner and compared it to the RSC-algorithm proposed by Sándor and Wedel (2001) to construct efficient choice designs. The designs under investigation were of type $3^4/2/15$. Since the so-called Bayesian modified Fedorov choice algorithm does not enforce any constraints on its design search as opposed to the RSC-algorithm, it provides the most efficient choice designs. To score the utility balances of the designs, we introduced as measure the percentage utility balance and found that the optimal designs are not utility balanced. Furthermore, we compared the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-optimality criteria and examined their estimation performance of the parameters and predictive validity. After all, conjoint choice experiments are carried out to predict the market demand of a specific product line or related services. We observed that the $\mathcal{A}_B$- and $\mathcal{D}_B$-optimal modified Fedorov designs provide the most accurate estimates and the $\mathcal{D}_B$- and $\mathcal{V}_B$-optimal modified Fedorov designs the most precise predictions. However, since the $\mathcal{D}_B$-optimality criterion results in the highest expected efficiency if the optimal design is missed and leads to the shortest computation time, it has to be preferred to the other optimality criteria to set up an effective choice experiment.

Acknowledgements

The research that lead to this paper was carried out while Peter Goos was a Postdoctoral Researcher of the Fund for Scientific Research – Flanders (Belgium). We would like to thank the faculty's IT department for putting some pc-rooms at our disposal and Dr. Yuri Goegebeur for his help with the graphical representations in S-PLUS 6.

Appendix A

To derive the information matrix $\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})$ in (4), we need to compute the Hessian matrix $\mathbf{H}(\mathbf{X}, \boldsymbol{\beta})$ since $\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})$ is defined as $-\mathrm{E}\{\mathbf{H}(\mathbf{X}, \boldsymbol{\beta})\}$. We start by calculating the gradient of the log-likelihood function from one respondent:

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} \ln \mathrm{L} &= \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js} \ln\{p_{js}(\mathbf{X_s}, \boldsymbol{\beta})\}, \\
&= \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js} \frac{\sum_{t=1}^{J} e^{\mathbf{x'_{ts}}\boldsymbol{\beta}}}{e^{\mathbf{x'_{js}}\boldsymbol{\beta}}} \left\{ \frac{e^{\mathbf{x'_{js}}\boldsymbol{\beta}} \mathbf{x_{js}} \left(\sum_{t=1}^{J} e^{\mathbf{x'_{ts}}\boldsymbol{\beta}} \right) - e^{\mathbf{x'_{js}}\boldsymbol{\beta}} \left(\sum_{t=1}^{J} e^{\mathbf{x'_{ts}}\boldsymbol{\beta}} \mathbf{x_{ts}} \right)}{\left(\sum_{t=1}^{J} e^{\mathbf{x'_{ts}}\boldsymbol{\beta}} \right)^2} \right\}, \\
&= \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js} \mathbf{x_{js}} - \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js} \frac{\sum_{t=1}^{J} e^{\mathbf{x'_{ts}}\boldsymbol{\beta}} \mathbf{x_{ts}}}{\sum_{t=1}^{J} e^{\mathbf{x'_{ts}}\boldsymbol{\beta}}}.
\end{aligned}
$$

The Hessian matrix $\mathbf{H}(\mathbf{X}, \boldsymbol{\beta})$ from one respondent then equals

$$\begin{aligned}
\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \ln \mathrm{L} &= \frac{\partial}{\partial\boldsymbol{\beta}'}\left(-\sum_{s=1}^{S}\sum_{j=1}^{J} y_{js}\frac{\sum_{t=1}^{J} e^{\mathbf{x}_{ts}'\boldsymbol{\beta}}\mathbf{x}_{ts}}{\sum_{t=1}^{J} e^{\mathbf{x}_{ts}'\boldsymbol{\beta}}}\right), \\
&= -\sum_{s=1}^{S}\sum_{j=1}^{J} y_{js}\left\{\frac{\sum_{t=1}^{J} e^{\mathbf{x}_{ts}'\boldsymbol{\beta}}\mathbf{x}_{ts}\mathbf{x}_{ts}' \sum_{t=1}^{J} e^{\mathbf{x}_{ts}'\boldsymbol{\beta}} - \sum_{t=1}^{J} e^{\mathbf{x}_{ts}'\boldsymbol{\beta}}\mathbf{x}_{ts} \sum_{t=1}^{J} e^{\mathbf{x}_{ts}'\boldsymbol{\beta}}\mathbf{x}_{ts}'}{\left(\sum_{t=1}^{J} e^{\mathbf{x}_{ts}'\boldsymbol{\beta}}\right)^2}\right\}, \\
&= -\sum_{s=1}^{S}\sum_{j=1}^{J} y_{js}\left(\sum_{t=1}^{J}\frac{e^{\mathbf{x}_{ts}'\boldsymbol{\beta}}}{\sum_{v=1}^{J} e^{\mathbf{x}_{vs}'\boldsymbol{\beta}}}\mathbf{x}_{ts}\mathbf{x}_{ts}'\right), \\
&\quad + \sum_{s=1}^{S}\sum_{j=1}^{J} y_{js}\left(\sum_{t=1}^{J}\frac{e^{\mathbf{x}_{ts}'\boldsymbol{\beta}}}{\sum_{v=1}^{J} e^{\mathbf{x}_{vs}'\boldsymbol{\beta}}}\mathbf{x}_{ts}\right)\left(\sum_{t=1}^{J}\frac{e^{\mathbf{x}_{ts}'\boldsymbol{\beta}}}{\sum_{v=1}^{J} e^{\mathbf{x}_{vs}'\boldsymbol{\beta}}}\mathbf{x}_{ts}'\right), \\
&= -\sum_{s=1}^{S}\sum_{j=1}^{J} y_{js}\left(\sum_{t=1}^{J} p_{ts}\mathbf{x}_{ts}\mathbf{x}_{ts}' - \sum_{t=1}^{J} p_{ts}\mathbf{x}_{ts}\sum_{t=1}^{J} p_{ts}\mathbf{x}_{ts}'\right), \\
&= -\sum_{s=1}^{S}\left(\sum_{t=1}^{J} p_{ts}\mathbf{x}_{ts}\mathbf{x}_{ts}' - \sum_{t=1}^{J} p_{ts}\mathbf{x}_{ts}\sum_{t=1}^{J} p_{ts}\mathbf{x}_{ts}'\right)\sum_{j=1}^{J} y_{js}, \\
&= -\sum_{s=1}^{S}\left(\sum_{t=1}^{J} p_{ts}\mathbf{x}_{ts}\mathbf{x}_{ts}' - \sum_{t=1}^{J} p_{ts}\mathbf{x}_{ts}\sum_{t=1}^{J} p_{ts}\mathbf{x}_{ts}'\right)1, \\
&= -\sum_{s=1}^{S}\left(\mathbf{X}_s'\mathbf{P}_s\mathbf{X}_s - \mathbf{X}_s'\mathbf{p}_s\mathbf{p}_s'\mathbf{X}_s\right), \\
&= -\sum_{s=1}^{S}\mathbf{X}_s'\left(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s'\right)\mathbf{X}_s,
\end{aligned}$$

where $\mathbf{X_s} = [\mathbf{x_{1s}}, ..., \mathbf{x_{Js}}]'$, $\mathbf{p_s} = [p_{1s}, ..., p_{Js}]'$ and $\mathbf{P_s} = \mathrm{diag}[p_{1s}, ..., p_{Js}]$. Taking the expected value of the Hessian matrix, $\mathrm{E}\{\mathbf{H}(\mathbf{X}, \boldsymbol{\beta})\}$, leads to the same expression:

$$\mathrm{E}\left(\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \ln \mathrm{L}\right) = -\sum_{s=1}^{S}\mathbf{X}_s'\left(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s'\right)\mathbf{X}_s.$$

As a result, the information matrix from N respondents becomes

$$\begin{aligned}
\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}) &= N\left(-\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \ln \mathrm{L}\right), \\
&= N\sum_{s=1}^{S}\mathbf{X}_s'\left(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s'\right)\mathbf{X}_s.
\end{aligned}$$

Appendix B

We compute the $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{G}_B$- and $\mathcal{V}_B$-criterion values of a small design $\mathbf{X^0}$, consisting of three choice sets each with two alternatives. The alternatives have two attributes: attribute 1 at three levels denoted 1, 2, 3, and coded [1 0], [0 1], [-1 -1] respectively, and attribute 2 at two levels, 1 and 2, coded -1 and 1. The design $\mathbf{X^0}$ with its coded version $\mathbf{X}$ looks as follows:

$$\mathbf{X^0} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ \hline 2 & 2 \\ 3 & 1 \\ \hline 3 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ \hline 0 & 1 & 1 \\ -1 & -1 & -1 \\ \hline -1 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

The three choice sets are separated by horizontal lines. The alternatives are specified in the rows and the columns represent the attributes. The Bayesian criterion values are computed as in (7), (10), (15) and (18). By way of illustration, we only use three prior parameters $\boldsymbol{\beta}^r = [\beta_{11}^r, \beta_{12}^r, \beta_2^r]$, $r = 1, 2, 3$, randomly drawn from $\pi(\boldsymbol{\beta}) = U[-1, 1]^k$ with $k = 3$. For each of these prior parameters, we compute the local $\mathcal{D}_P^r$-, $\mathcal{A}_P^r$-, $\mathcal{G}_P^r$- and $\mathcal{V}_P^r$-criterion values and subsequently average them to obtain the corresponding Bayesian values.

The information matrix $\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^r)$ is computed as in (4) by taking $N = 1$ so that $\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^r) = \sum_{s=1}^{3} \mathbf{X_s'}(\mathbf{P_s} - \mathbf{p_s p_s'})\mathbf{X_s}$ with choice sets given by

$$\mathbf{X_1} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}, \mathbf{X_2} = \begin{bmatrix} 0 & 1 & 1 \\ -1 & -1 & -1 \end{bmatrix}, \mathbf{X_3} = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix},$$

$\mathbf{p_s}$ the vector of probabilities in choice set s and $\mathbf{P_s}$ the corresponding diagonal matrix.

As a first draw, $r = 1$, we have $\boldsymbol{\beta}^1 = [-0.238, 0.656, 0.122]'$. Using (2), we obtain the following for choice set $s = 1$:

$$\mathbf{p_1} = \begin{bmatrix} 0.343 \\ 0.657 \end{bmatrix}, \mathbf{P_1} - \mathbf{p_1 p_1'} = 0.225 \times \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \text{ and}$$

$$\mathbf{X_1'}(\mathbf{P_1} - \mathbf{p_1 p_1'})\mathbf{X_1} = \begin{bmatrix} 0.225 & -0.225 & 0.451 \\ -0.225 & 0.225 & -0.451 \\ 0.451 & -0.451 & 0.901 \end{bmatrix}.$$

Repeating the computations for choice sets 2 and 3 and summing the three matrices yields the information matrix corresponding to $\boldsymbol{\beta}^1$:

$$\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1) = \begin{bmatrix} 1.391 & 0.607 & -0.215 \\ 0.607 & 1.141 & -0.284 \\ -0.215 & -0.284 & 2.567 \end{bmatrix}.$$

The local $\mathcal{D}_P^1$-criterion value then becomes

$$\mathcal{D}_P^1 = \{\text{det}\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\}^{1/3} = 0.691,$$

and the local $\mathcal{A}_P^1$-criterion value,

$$\mathcal{A}_P^1 = \text{tr}\{\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\} = 2.499.$$

To obtain the local $\mathcal{G}_P^1$- and $\mathcal{V}_P^1$-criterion values, we use the more practical design region $\chi*$ that consists of one large choice set embracing all candidate alternatives. For our small example, since attribute 1 has three levels and attribute 2 has two levels, this large choice set, C_l, contains $3 \times 2 = 6$ alternatives: $C_l = \{\mathbf{x_{1l}}, ..., \mathbf{x_{6l}}\}$. For each of these alternatives, we compute the corresponding $\mathbf{c}$-vector according to (12) and (13):

$$C_l = \begin{array}{|ccc|} \hline 1 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \\ -1 & -1 & -1 \\ -1 & -1 & 1 \\ \hline \end{array} \begin{array}{ccc} \rightarrow & p_{1l} = 0.103 & \rightarrow \quad \mathbf{c}(\mathbf{x_{1l}}) = [0.099, -0.039, -0.115]', \\ \rightarrow & p_{2l} = 0.131 & \rightarrow \quad \mathbf{c}(\mathbf{x_{2l}}) = [0.126, -0.049, 0.115]', \\ \rightarrow & p_{3l} = 0.251 & \rightarrow \quad \mathbf{c}(\mathbf{x_{3l}}) = [-0.010, 0.157, -0.281]', \\ \rightarrow & p_{4l} = 0.320 & \rightarrow \quad \mathbf{c}(\mathbf{x_{4l}}) = [-0.012, 0.200, 0.281]', \\ \rightarrow & p_{5l} = 0.086 & \rightarrow \quad \mathbf{c}(\mathbf{x_{5l}}) = [-0.089, -0.118, -0.096]', \\ \rightarrow & p_{6l} = 0.109 & \rightarrow \quad \mathbf{c}(\mathbf{x_{6l}}) = [-0.114, -0.151, 0.096]'. \end{array}$$

The local $\mathcal{G}_P^1$- and $\mathcal{V}_P^1$-criterion values are then derived as

$$\mathcal{G}_P^1 = \max \begin{pmatrix} \mathbf{c}'(\mathbf{x_{1l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{1l}}) = 0.020 \\ \mathbf{c}'(\mathbf{x_{2l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{2l}}) = 0.029 \\ \mathbf{c}'(\mathbf{x_{3l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{3l}}) = 0.054 \\ \mathbf{c}'(\mathbf{x_{4l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{4l}}) = 0.090 \\ \mathbf{c}'(\mathbf{x_{5l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{5l}}) = 0.019 \\ \mathbf{c}'(\mathbf{x_{6l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{6l}}) = 0.022 \end{pmatrix} = 0.090,$$

$$\mathcal{V}_P^1 = \text{avg} \begin{pmatrix} \mathbf{c}'(\mathbf{x_{1l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{1l}}) = 0.020 \\ \mathbf{c}'(\mathbf{x_{2l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{2l}}) = 0.029 \\ \mathbf{c}'(\mathbf{x_{3l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{3l}}) = 0.054 \\ \mathbf{c}'(\mathbf{x_{4l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{4l}}) = 0.090 \\ \mathbf{c}'(\mathbf{x_{5l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{5l}}) = 0.019 \\ \mathbf{c}'(\mathbf{x_{6l}})\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}^1)^{-1}\mathbf{c}(\mathbf{x_{6l}}) = 0.022 \end{pmatrix} = 0.039.$$

Similar computations for a second draw, $r = 2$, $\boldsymbol{\beta}^2 = [0.045, -0.832, -0.198]'$, yield

$$\begin{aligned} \mathcal{D}_P^2 &= 0.804, \\ \mathcal{A}_P^2 &= 2.969, \\ \mathcal{G}_P^2 &= 0.140, \\ \mathcal{V}_P^2 &= 0.048, \end{aligned}$$

and for a third draw, $r = 3$, $\boldsymbol{\beta}^3 = [0.783, -0.267, 0.549]'$, we obtain

$$\begin{aligned}\mathcal{D}_P^3 &= 0.934,\\ \mathcal{A}_P^3 &= 4.080,\\ \mathcal{G}_P^3 &= 0.198,\\ \mathcal{V}_P^3 &= 0.058.\end{aligned}$$

Finally, we average the local criterion values over the three draws to get the corresponding Bayesian values:

$$\begin{aligned}\widetilde{\mathcal{D}}_B(\mathbf{X^0}) &= 0.809,\\ \widetilde{\mathcal{A}}_B(\mathbf{X^0}) &= 3.183,\\ \widetilde{\mathcal{G}}_B(\mathbf{X^0}) &= 0.143,\\ \widetilde{\mathcal{V}}_B(\mathbf{X^0}) &= 0.048.\end{aligned}$$

References

ACA system (Version 4.0)[Computer software](1994). Evanston, IL: Sawtooth Software.

Arora, N. and Huber, J. (2001). Improving parameter estimates and model prediction by aggregate customization in choice experiments. *Journal of Consumer Research* **28**: 273–283.

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*, Oxford U.K.: Clarendon Press.

Atkinson, A. C. and Haines, L. M. (1996). Designs for nonlinear and generalized linear models, in *Handbook of Statistics 13: Design and Analysis of Experiments*, composed by Gosh, S. and Rao, C., Elsevier, Amsterdam, 437–475.

Bunch, D. S., Louviere, J. J. and Anderson, D. A. (1994). A comparison of experimental design strategies for multinomial logit models: the case of generic attributes, working paper, Graduate School of Management, University of California at Davis.

Carson, R. T., Louviere, J. J., Anderson, D. A., Arabie, P., Bunch, D. S., Hensher, D. A., Johnson R. M., Kuhfeld W. F., Steinberg D., Swait, J., Timmermans, H. and Wiley, J. B. (1994). Experimental analysis of choice, *Marketing Letters* **5**: 351–367.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: a review, *Statistical Science* **10**, 273–304.

Cook, R. D. and Nachtsheim, C. J. (1980). A comparison of algorithms for constructing exact $\mathcal{D}$-optimal designs, *Technometrics* **22**: 315–324.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*, New York: Academic Press.

Goos, P. (2002). *The Optimal Design of Blocked and Split-plot Experiments*, New York: Springer.

Grossmann, H., Holling H. and Schwabe R. (2002). Advances in optimum experimental design for conjoint analysis and discrete choice models, *Econometric Models in Marketing, Vol. 16*, 93–117.

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. (2002). Optimal design for multinomial choice experiments, *Journal of Marketing Research* **39**: 214–227.

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

Lazari, A. G. and Anderson, D. (1994). Designs of discrete choice experiments for estimating both attribute and availability cross effects, *Journal of Marketing Research* **31**: 375–383.

Louviere, J. J. and Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data, *Journal of Marketing Research* **20**: 350–367.

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

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**: 55–75.

Sándor, Z. and Wedel, M. (2003). Differentiated Bayesian conjoint choice designs, Erasmus Research Institute of Management (ERIM)–report ERS–2003–016–MKT, Erasmus University Rotterdam, The Netherlands.

Silvey, S. D. (1980). *Optimal Design*, London: Chapman and Hall.

Trinca, L. A. and Gilmour, S. G. (2000). An algorithm for arranging response surface designs in small blocks, *Computational Statistics and Data Analysis* **33**: 25–43.

Van Lenthe, J. (1993). ELI: An interactive elicitation technique for subjective probability distributions, *Organizational Behavior and Human Decision Processes* **55**: 379–413.

Zwerina, K., Huber, J. and Kuhfeld, W. F. (1996). A general method for constructing efficient choice designs, working paper, Fuqua School of Business, Duke University, Durham, NC 27708. Available from SAS Institute (http://ftp.sas.com/techsup/download/technote/ts629.pdf).