DEPARTMENT OF MATHEMATICS,
STATISTICS AND ACTUARIAL SCIENCES

D-optimal conjoint choice designs with no-choice options for a nested logit model

Peter Goos, Bart Vermeulen & Martina Vandebroek

UNIVERSITY OF ANTWERP
Faculty of Applied Economics

Stadscampus
Prinsstraat 13, B.213
BE-2000 Antwerpen
Tel. +32 (0)3 220 40 32
Fax +32 (0)3 220 47 99
http://www.ua.ac.be/tew

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF MATHEMATICS,
STATISTICS AND ACTUARIAL SCIENCES

D-optimal conjoint choice designs with no-choice options for a nested logit model

Peter Goos, Bart Vermeulen & Martina Vandebroek

RESEARCH PAPER 2008-020
DECEMBER 2008

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.213
phone: (32) 3 220 40 32
fax: (32) 3 220 47 99
e-mail: joeri.nys@ua.ac.be

D/2008/1169/020

D-optimal conjoint choice designs with no-choice options for a nested logit model

Peter Goos[1]

Universiteit Antwerpen

Bart Vermeulen, Martina Vandebroek

Katholieke Universiteit Leuven

Abstract

Despite the fact that many conjoint choice experiments offer respondents a no-choice option in every choice set, the optimal design of conjoint choice experiments involving no-choice options has received only a limited amount of attention in the literature. In this article, we present an approach to construct D-optimal designs for this type of experiment. For that purpose, we derive the information matrix of a nested multinomial logit model that is appropriate for analyzing data from choice experiments with no-choice options. The newly derived information matrix is compared to the information matrix for the multinomial logit model that is used in the literature to construct designs for choice experiments. It is also used to quantify the loss of information in a choice experiment due to the presence of a no-choice option.

Keywords: choice-based conjoint, information loss, multinomial logit model, nested logit model, no-choice option

1 Introduction

Conjoint choice experimentation, in which respondents indicate the alternative they like most in one or more sets of alternatives, is commonly used in marketing for quantifying consumers' preferences for the attributes of products and services. The design of conjoint choice experiments that allow an efficient measurement of these preferences has therefore received a substantial amount of attention in the marketing literature and in the statistical literature. Most of the recent approaches apply the principles of the optimal design of experiments (see, e.g., Atkinson et al. 2007). Theoretical results concerning optimal designs for paired comparison experiments, which are choice experiments involving two alternatives per choice set, were obtained by Burgess and Street (2003, 2005), Grasshoff et al. (2003, 2004), Grossmann et al. (2002, 2006), Street et al. (2001), Street and Burgess (2004a), van Berkum (1987) and van Berkum and Pauwels (2003). These results are based on the a priori assumption that respondents are indifferent between each of the alternatives offered. This assumption, which is often criticized for not being realistic

[1]Corresponding author. Postal address: Peter Goos, Faculteit Toegepaste Economische Wetenschappen, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium. Tel ++32 3 220 4059. Fax ++32 3 220 48 17. Email peter.goos@ua.ac.be

and leading to non-informative choice sets, reduces the optimal design problem for a nonlinear statistical model to the simpler problem of finding an optimal block design with blocks of two observations for a linear model. Computational approaches to seek optimal designs for conjoint choice experiments with any number of alternatives per choice set can be found in Huber and Zwerina (1996), Sándor and Wedel (2001, 2002, 2005) and Kessels et al. (2006, 2008a, 2008b), neither of whom make the simplifying assumption of indifference between the alternatives. Kanninen (2002) derives optimal choice designs for situations involving continuous attributes one of which is unrestricted, and Grasshoff et al. (2007) obtain results for choice experiments with two alternatives per choice set and one unrestricted quantitative attribute. In all of these contributions, the focus is on choice experiments in which the respondents are forced to indicate their preference for one of the alternatives in a choice set. This is remarkable as many choice experiments leave the respondents the option not to choose if none of the alternatives in a choice set pleases them. As a matter of fact, the choice sets in many experiments do not only include several alternatives, but they also include an option which is common to every choice set. That option, sometimes called the base alternative, is often labelled as "None of these" or "I would stick to my current brand". We refer to this option, which is common to all choice sets in the experiment, as the no-choice option.

The optimal design of choice experiments including such a no-choice option received attention by Street and Burgess (2004b), who discovered that, apart from the no-choice option in every choice set and under the a priori assumption that the respondents are indifferent between the alternatives, the optimal designs for experiments with and without no-choice option are equal. In this article, we focus on the optimal design of choice experiments with a no-choice option and do not use the simplifying indifference assumption. Instead of using the multinomial logit model utilized as a basis for the computational approaches by Huber and Zwerina (1996), Sándor and Wedel (2001, 2005) and Kessels et al. (2006, 2008a), we use the nested multinomial logit model as a basis for computing the optimal designs. This model is the most appealing and most advanced one of three possible models discussed by Haaijer et al. (2001) for analyzing data from experiments with a no-choice option. The model implicitly assumes that respondents go through a two-stage decision process when evaluating a choice set containing a number of alternatives, henceforth called real-choice options, and a no-choice option. First, they compare the real-choice options and select the most attractive one. Next, the respondents decide whether or not the utility of that alternative is sufficiently high to choose it instead of the no-choice option.

In this article, we derive the information matrix corresponding to the nested logit model for analyzing data from choice experiments with a no-choice option. This allows us to formulate the D-optimality criterion and a D_s-optimality criterion for seeking designs that allow an efficient estimation of the parameters of the nested logit model. A theoretical result is derived and an example is given to show the impact on the optimal conjoint choice designs of taking into account the no-choice option by using the nested logit model. Also, it is shown how the loss of information due to the no-choice option can be measured. First however, we briefly review the multinomial logit model and the resulting D-optimality criterion, as well as the general nested multinomial logit model. For an in-depth discussion of the multinomial logit model and the nested logit model, we refer the reader to Train (2003).

2 Optimal designs for the multinomial logit model

The multinomial logit model for analyzing choice data assumes that the utility a respondent attaches to the jth alternative in the ith choice set can be modelled as

$$U_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \varepsilon_{ij},$$

where $\mathbf{x}_{ij}$ is a q-dimensional vector containing the model expansion of the attribute levels of the jth alternative in choice set i, $\boldsymbol{\beta}$ contains the q model coefficients that describe the respondents' preferences and ε_{ij} is a random error term. It is assumed that all error terms are independent and identically Gumbel distributed. As a result, the probability that alternative j is chosen in choice set i can be written as

$$p^*_{ij} = \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}}{\sum_{k=1}^{J} e^{\mathbf{x}'_{ik}\boldsymbol{\beta}}},$$

where J denotes the number of alternatives in the choice set. The multinomial logit model is estimated by maximizing the likelihood function, which, for choice set i and a single respondent, is given by

$$L^*_i = \prod_{j=1}^{J} (p^*_{ij})^{y_{ij}},$$

where y_{ij} is equal to one if the respondent chooses alternative j in choice set i and zero otherwise. This expression can be used to derive the information matrix on the unknown model parameter $\boldsymbol{\beta}$ for choice set i:

$$\mathbf{M}^*_i = \mathbf{X}'_i(\mathbf{P}^*_i - \mathbf{p}^*_i\mathbf{p}^{*'}_i)\mathbf{X}_i,$$

where $\mathbf{X}_i = [\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iJ}]'$, $\mathbf{P}^*_i = \text{diag}(p^*_{i1}, \ldots, p^*_{iJ})$ and $\mathbf{p}^*_i = [p^*_{i1}, \ldots, p^*_{iJ}]'$. The total information matrix is

$$\mathbf{M}^* = \sum_{i=1}^{S} \mathbf{M}^*_i = \mathbf{X}'\mathbf{D}^*\mathbf{X}, \tag{1}$$

where S is the number of choice sets in the experiment, $\mathbf{X} = [\mathbf{X}'_1, \ldots, \mathbf{X}'_S]'$, $\mathbf{D}^* = \text{diag}(\mathbf{D}^*_1, \ldots, \mathbf{D}^*_S)$ and $\mathbf{D}^*_i = \mathbf{P}^*_i - \mathbf{p}^*_i\mathbf{p}^{*'}_i$.

This information matrix forms the basis for the computation of optimal designs for conjoint choice experiments in Huber and Zwerina (1996), Sándor and Wedel (2001, 2005) and Kessels et al. (2006, 2008a). As the information matrix depends on $\boldsymbol{\beta}$ through the probabilities p^*_{ij} in $\mathbf{P}^*_i$, $\mathbf{p}^*_i$ and $\mathbf{D}^*$, the optimal designs also depend on it. This is typical for nonlinear models and implies that prior information on the unknown $\boldsymbol{\beta}$ is required to construct optimal conjoint choice designs. Huber and Zwerina (1996) used a point estimate for $\boldsymbol{\beta}$ as an input to their search procedure for optimal conjoint choice designs, whereas Sándor and Wedel (2001, 2005) and Kessels et al. (2006, 2008a) use a prior distribution for $\boldsymbol{\beta}$. The optimal designs obtained by Huber and Zwerina (1996) are usually referred to as locally optimal designs, whereas those obtained by Sándor and Wedel (2001, 2005) and Kessels et al. (2006, 2008a) are called Bayesian optimal designs. The optimal designs presented by Burgess and Street (2003, 2005), Grasshoff et al. (2003, 2004), Grossmann et al. (2002, 2006), Street et al. (2001), Street and Burgess (2004a), van Berkum (1987) (1987) and van Berkum and Pauwels (2003) were constructed assuming $\boldsymbol{\beta} = \mathbf{0}_q$ and can therefore be seen as locally optimal designs. The assumption that $\boldsymbol{\beta} = \mathbf{0}_q$, where $\mathbf{0}_q$ is a q-dimensional vector of zeroes, causes the probabilities p^*_{ij} of all J alternatives in choice set i to be equal to $1/J$, which corresponds to a situation in which respondents have no preference for any of the alternatives in the choice set. The information matrix for choice set i then simplifies to

$$\mathbf{M}^{0*}_i = J^{-1}\mathbf{X}'_i(\mathbf{I}_J - J^{-1}\mathbf{1}_J\mathbf{1}'_J)\mathbf{X}_i, \tag{2}$$

where $\mathbf{I}_J$ and $\mathbf{1}_J$ denote the J-dimensional identity matrix and a J-dimensional vector of ones, respectively. This is because, when all $p^*_{ij} = 1/J$, $\mathbf{P}^*_i = J^{-1}\mathbf{I}_J$ and $\mathbf{p}^*_i = J^{-1}\mathbf{1}_J$. The total information matrix then equals

$$\mathbf{M}^{0*} = \sum_{i=1}^{S} \mathbf{M}^{0*}_i = J^{-1}\{\mathbf{X}'\mathbf{X} - \sum_{i=1}^{S} J^{-1}(\mathbf{X}'_i\mathbf{1}_J)(\mathbf{1}'_J\mathbf{X}_i)\}, \tag{3}$$

which, up to a proportionality constant, is equal to the information matrix on $\boldsymbol{\beta}$ from a blocked experiment with blocks of J observations and where the block effects are treated as fixed parameters (see, e.g., Goos (2002)). As a result, locally optimal designs obtained assuming $\boldsymbol{\beta} = \mathbf{0}_q$ are exactly the same as optimal designs for blocked experiments when the block effects are treated as fixed parameters.

3 Nested logit model

The multinomial logit model assumes that all error terms ε_{ij} are independent from each other. Frequently however, this assumption, which is referred to as the independence of irrelevant alternatives assumption, is invalid because two or more alternatives in a choice set are similar so that their utilities U_{ij} are correlated. In such cases, the nested multinomial logit model, which partitions the J alternatives in a choice set in M nests $\mathcal{J}_1, \ldots, \mathcal{J}_M$ of sizes $J_1, \ldots, J_M$, is often appropriate. Each nest consists of similar alternatives, i.e. alternatives whose utilities are correlated. The rationale behind the nested logit model is that respondents go through a two-stage decision process. First, they select the best alternative in each nest, and second, they select the nest that has the most attractive alternative selected in the first stage. This decision process is modelled by combining two ordinary multinomial logit models. The probability that nest m is chosen from the set of M nests in choice set i is modelled as

$$p_{i,m} = \frac{e^{\lambda_m V_{i,m}}}{\sum_{k=1}^{M} e^{\lambda_k V_{i,k}}},$$

where

$$V_{i,m} = \log\Big(\sum_{j \in \mathcal{J}_m} e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}\Big)$$

is the so-called inclusive value of the mth nest in choice set i and λ_m is its dissimilarity coefficient. The inclusive value $V_{i,m}$ is a measure for the utility of nest m, whereas λ_m expresses to what extent the alternatives within nest m are different. Most often, each λ_m takes values between zero and one. If all λ_m equal one, then the nested logit model reduces to the ordinary multinomial logit model. In general, the conditional probability that alternative j is chosen within nest m, given that nest m is the preferred nest in choice set i, is modelled as

$$p_{ij|m} = \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}}{\sum_{k \in \mathcal{J}_m} e^{\mathbf{x}'_{ik}\boldsymbol{\beta}}}.$$

The probability that alternative j, which belongs to nest m, is chosen from choice set i equals $\overline{p}_{ij} = p_{ij|m} p_{i,m}$. The likelihood function for choice set i for the nested logit model can then be written as

$$\overline{L}_i = \prod_{j=1}^{J} \overline{p}_{ij}^{\,y_{ij}},$$

where y_{ij} is again equal to one if the respondent chooses alternative j in choice set i and zero otherwise.

4 No-choice nested logit model

For analyzing stated preference data from a choice experiment with a no-choice option, Haaijer et al. (2001) suggest using two nests in the nested logit model. The first nest contains J real-choice options and the second nest contains the no-choice option only. We refer to the resulting model as the no-choice nested logit model.

4.1 Model

For choice set i, we denote the probability that the jth real-choice alternative is chosen by p_{ij}, the probability that a real-choice option is chosen by $p_{i,C}$ and the probability that the no-choice option is selected by $p_{i,NC}$. The probability that a respondent selects a real-choice option is then modelled as

$$p_{i,C} = \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{NC} V_{i,NC}}},$$

whereas the probability that the nest with the no-choice option is chosen is modelled as

$$p_{i,NC} = \frac{e^{\lambda_{NC} V_{i,NC}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{NC} V_{i,NC}}}.$$

In these expressions, λ_C and λ_{NC} are the dissimilarity parameters for the nest with real-choice options and the nest with the no-choice option, and $V_{i,C}$ and $V_{i,NC}$ represent the inclusive values of the nest with the real-choice options and the nest with the no-choice option, respectively. The former inclusive value is equal to

$$V_{i,C} = \log\Big(\sum_{j=1}^{J} e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}\Big).$$

The inclusive value for the nest with the no-choice option can be simplified because it contains only one option:

$$V_{i,NC} = \log\big(e^{\mathbf{x}'_{i,NC}\boldsymbol{\beta}}\big) = \mathbf{x}'_{i,NC}\boldsymbol{\beta}.$$

In this article, we follow the suggestion of Haaijer et al. (2001) to use $\mathbf{x}_{i,NC} = \mathbf{0}_q$ so that $V_{i,NC}$ is simply zero. As a result,

$$p_{i,C} = \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + 1}$$

and

$$p_{i,NC} = \frac{1}{e^{\lambda_C V_{i,C}} + 1}.$$

The probability that the jth real-choice alternative is chosen is equal to

$$p_{ij} = p_{i,C} \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}}{\sum_{k=1}^{J} e^{\mathbf{x}'_{ik}\boldsymbol{\beta}}}.$$

It is not hard to see that $\sum_{j=1}^{J} p_{ij} = p_{i,C} = 1 - p_{i,NC}$. For notational convenience, we will henceforth drop the subscript C from λ_C and denote the dissimilarity parameter of the nest with real-choice options simply by λ.

A key feature of the no-choice nested logit model is the fact that the utility of the no-choice option is assumed to be zero, by setting $\mathbf{x}_{i,NC} = \mathbf{0}_q$. While this raises no concerns when all the attributes are treated as categorical (as in most work on the design of conjoint choice experiments), it may seem inappropriate in the presence of attributes that are treated as quantitative. It is true that such an assumption would lead to biased estimates for the parameters of the quantitative attributes in many models. In the no-choice nested logit model, however, the no-choice option is in a separate nest which implies that it is modelled as an alternative that cannot be compared directly to the real-choice options. Almost all the information for estimating the parameters contained within $\boldsymbol{\beta}$ appears in the nest with the real-choice options. This implies that $\boldsymbol{\beta}$ is estimated with information acquired from respondents that choose a real-choice option. As a result, there is no need for a concern about bias of the parameter estimates.

4.2 Loglikelihood function

Estimating the no-choice nested logit model can be done by maximizing the likelihood function with respect to the unknown parameters $\boldsymbol{\beta}$ and λ. To write down the likelihood function and the loglikelihood function for the no-choice nested logit model, let $y_{i,NC}$ be one if the no-choice option is selected in choice set i and zero otherwise, $y_{i,C}$ be one if a real-choice option is selected and zero otherwise, and y_{ij} be one if the jth real-choice option is selected and zero otherwise. Evidently, $y_{i,C} = \sum_{j=1}^{J} y_{ij}$ and $y_{i,NC} + \sum_{j=1}^{J} y_{ij} = 1$. The likelihood function for choice set i can then be written as

$$
\begin{aligned}
L_i &= p_{i,NC}^{y_{i,NC}} \prod_{j=1}^{J} p_{ij}^{y_{ij}}, \\
&= \Big(\frac{1}{e^{\lambda V_{i,C}} + 1}\Big)^{y_{i,NC}} \prod_{j=1}^{J} \Big(\frac{e^{\lambda V_{i,C}}}{e^{\lambda V_{i,C}} + 1} \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}}{\sum_{k=1}^{J} e^{\mathbf{x}'_{ik}\boldsymbol{\beta}}}\Big)^{y_{ij}}, \\
&= \frac{e^{\lambda V_{i,C} \sum_{j=1}^{J} y_{ij}}}{(e^{\lambda V_{i,C}} + 1)^{y_{i,NC} + \sum_{j=1}^{J} y_{ij}} \, (\sum_{k=1}^{J} e^{\mathbf{x}'_{ik}\boldsymbol{\beta}})^{\sum_{j=1}^{J} y_{ij}}} \prod_{j=1}^{J} e^{y_{ij}\mathbf{x}'_{ij}\boldsymbol{\beta}}, \\
&= \frac{e^{y_{i,C}\lambda V_{i,C}}}{(e^{\lambda V_{i,C}} + 1)(\sum_{k=1}^{J} e^{\mathbf{x}'_{ik}\boldsymbol{\beta}})^{y_{i,C}}} \prod_{j=1}^{J} e^{y_{ij}\mathbf{x}'_{ij}\boldsymbol{\beta}},
\end{aligned}
$$

so that the loglikelihood function is

$$
\begin{aligned}
l_i &= y_{i,C}\lambda V_{i,C} - \log(e^{\lambda V_{i,C}} + 1) - y_{i,C} \, \log(\sum_{k=1}^{J} e^{\mathbf{x}'_{ik}\boldsymbol{\beta}}) + \sum_{j=1}^{J} y_{ij}\mathbf{x}'_{ij}\boldsymbol{\beta}, \\
&= y_{i,C}(\lambda - 1)V_{i,C} - \log(e^{\lambda V_{i,C}} + 1) + \sum_{j=1}^{J} y_{ij}\mathbf{x}'_{ij}\boldsymbol{\beta}.
\end{aligned}
$$

4.3 Information matrix

From the rather simple expression for the loglikelihood function, the information matrix on the unknown model parameters $\boldsymbol{\beta}$ and λ can be derived. In Appendix A, it is shown that the information matrix for choice set i can be written as

$$
\mathbf{M}_i = \begin{bmatrix} \mathbf{X}'_i\mathbf{D}_i\mathbf{X}_i & p_{i,NC}\lambda V_{i,C}\mathbf{X}'_i\mathbf{p}_i \\ p_{i,NC}\lambda V_{i,C}\mathbf{p}'_i\mathbf{X}_i & p_{i,C}p_{i,NC}V_{i,C}^2 \end{bmatrix},
$$

where $\mathbf{X}_i = [\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iJ}]'$, $\mathbf{D}_i = [\, \mathbf{P}_i + p_{i,C}^{-1}\{\lambda^2 p_{i,NC} - 1\}\mathbf{p}_i\mathbf{p}'_i \,]$, $\mathbf{P}_i = \text{diag}(p_{i1}, \ldots, p_{iJ})$ and $\mathbf{p}_i = [p_{i1}, \ldots, p_{iJ}]'$. The upper left hand part of $\mathbf{M}_i$, $\mathbf{X}'_i\mathbf{D}_i\mathbf{X}_i$, is equal to the information matrix for the multinomial logit model, $\mathbf{M}_i^*$, if $p_{i,NC} = 0$ and thus $p_{i,C} = 1$.

If we again denote the number of choice sets in the experiment by S, the total information matrix for the no-choice nested logit model can be written as

$$\mathbf{M} = \sum_{i=1}^{S} \mathbf{M}_i = \begin{bmatrix} \sum_{i=1}^{S} \mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i & \sum_{i=1}^{S} p_{i,NC}\lambda V_{i,C}\mathbf{X}_i'\mathbf{p}_i \\ \sum_{i=1}^{S} p_{i,NC}\lambda V_{i,C}\mathbf{p}_i'\mathbf{X}_i & \sum_{i=1}^{S} p_{i,C}p_{i,NC}V_{i,C}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{D}\mathbf{X} & \mathbf{X}'\boldsymbol{\pi} \\ \boldsymbol{\pi}'\mathbf{X} & t \end{bmatrix}, \quad (4)$$

where $\mathbf{X} = [\mathbf{X}_1', \ldots, \mathbf{X}_S']'$, $\mathbf{D} = \text{diag}(\mathbf{D}_1, \ldots, \mathbf{D}_S)$, $t = \sum_{i=1}^{S} p_{i,C}p_{i,NC}V_{i,C}^2$ and $\boldsymbol{\pi} = \lambda[p_{1,NC}V_{1,C}\mathbf{p}_1', \ldots, p_{S,NC}V_{S,C}\mathbf{p}_S']'$.

The per choice set information matrix and the total information matrix depend on the unknown model parameters $\boldsymbol{\beta}$ and λ through $\mathbf{D}_i$, $p_{i,NC}$, $p_{i,C}$, $V_{i,C}$ and $\mathbf{p}_i$, and through $\mathbf{D}$, c and $\boldsymbol{\pi}$, respectively.

If the simplifying assumption that $\boldsymbol{\beta} = \mathbf{0}_q$ is made, then $V_{i,C} = \log J$, $p_{i,C} = J^\lambda/(J^\lambda + 1)$, $p_{i,NC} = 1/(J^\lambda + 1)$ and $p_{ij} = J^{\lambda-1}/(J^\lambda + 1)$. The information matrix for choice set i can then be written as

$$\mathbf{M}_i^0 = \frac{J^{\lambda-1}}{J^\lambda + 1} \begin{bmatrix} \mathbf{X}_i'(\mathbf{I}_J + J^{-1}\{\lambda^2(J^\lambda + 1)^{-1} - 1\}\mathbf{1}_J\mathbf{1}_J')\mathbf{X}_i & \lambda(J^\lambda + 1)^{-1}(\log J)\mathbf{X}_i'\mathbf{1}_J \\ \lambda(J^\lambda + 1)^{-1}(\log J)\mathbf{1}_J'\mathbf{X}_i & J(J^\lambda + 1)^{-1}\log^2 J \end{bmatrix}.$$

The total information matrix, $\mathbf{M}^0 = \sum_{i=1}^{S} \mathbf{M}_i^0$, then becomes

$$\mathbf{M}^0 = \frac{J^{\lambda-1}}{J^\lambda + 1} \begin{bmatrix} \mathbf{X}'\mathbf{X} + J^{-1}\{\lambda^2(J^\lambda + 1)^{-1} - 1\}\sum_{i=1}^{S} \mathbf{X}_i'\mathbf{1}_J\mathbf{1}_J'\mathbf{X}_i & \lambda(J^\lambda + 1)^{-1}(\log J)\mathbf{X}'\mathbf{1}_n \\ \lambda(J^\lambda + 1)^{-1}(\log J)\mathbf{1}_n'\mathbf{X} & n(J^\lambda + 1)^{-1}\log^2 J \end{bmatrix},$$

where $n = SJ$ is the total number of real-choice options in the conjoint choice design.

4.4 Design criteria

The goal of performing a choice experiment is usually to estimate the parameter vector $\boldsymbol{\beta}$ as precisely as possible, while precise inference about λ is not important. The researcher's primary interest is thus in a subset of the model parameters, and this has to be reflected in the criterion for selecting a design for the choice experiment. An appropriate criterion for selecting a design in such cases is the D_s-optimality criterion which seeks designs that minimize the generalized variance of a subset of the model parameters (see, for example, Atkinson et al. (2007)). The generalized variance of the parameters of interest here is given by the determinant of the part of the parameter estimates' variance-covariance matrix corresponding to $\boldsymbol{\beta}$. As this variance-covariance matrix is the inverse of the information matrix in (4), the D_s-optimal design is obtained by minimizing

$$\det(\mathbf{X}'\mathbf{D}\mathbf{X} - t^{-1}\mathbf{X}'\boldsymbol{\pi}\boldsymbol{\pi}'\mathbf{X}) = \{1 - t^{-1}\boldsymbol{\pi}'\mathbf{X}(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\pi}\}\det(\mathbf{X}'\mathbf{D}\mathbf{X}), \quad (5)$$

which we name the $D_{\boldsymbol{\beta}}$-optimality criterion.

If the simplifying assumption that $\boldsymbol{\beta} = \mathbf{0}_q$ is made, then the $D_{\boldsymbol{\beta}}$-optimality criterion can be written as

$$\left(\frac{J^{\lambda-1}}{J^\lambda+1}\right)^q \det[\,\mathbf{X}'\mathbf{X}+J^{-1}\{\lambda^2(J^\lambda+1)^{-1}-1\}\sum_{i=1}^{S}\mathbf{X}_i'\mathbf{1}_J\mathbf{1}_J'\mathbf{X}_i-n^{-1}\lambda^2J^{\lambda-1}(J^\lambda+1)^{-2}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\,]. \tag{6}$$

This expression reduces to

$$\det[\,(J+1)^{-1}\{\mathbf{X}'\mathbf{X}-(J+1)^{-1}\sum_{i=1}^{S}(\mathbf{X}_i'\mathbf{1}_J)(\mathbf{1}_J'\mathbf{X}_i)\}-n^{-1}(J+1)^{-2}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\,] \tag{7}$$

when $\lambda = 1$. Apart from the last term, the matrix in this expression is similar to the information matrix in (3). Now, for most sensible experimental designs, $\mathbf{X}'\mathbf{1}_n$ is equal to or close to the zero vector when effects-type coding is used, meaning that, for each attribute, each level appears equally often or almost equally often in the design. A consequence of this is that, under the assumptions that $\boldsymbol{\beta} = \mathbf{0}_q$ and $\lambda = 1$, the design optimality criteria for the multinomial logit model in Section 2 and for the no-choice nested logit model are nearly equal. As a result, designs that are optimal for the multinomial logit model assuming that $\boldsymbol{\beta} = \mathbf{0}_q$ will be at least nearly-optimal for the no-choice nested logit model too when $\boldsymbol{\beta} = \mathbf{0}_q$ and λ is close to one.

If the researcher happens to be interested in a precise estimation of both $\boldsymbol{\beta}$ and λ, the usual D-optimality criterion, which seeks designs that maximize the determinant of the full information matrix, can be used. The D-optimality criterion in that case is

$$\det(\mathbf{M}) = t\{1 - t^{-1}\boldsymbol{\pi}'\mathbf{X}(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\pi}\}\det(\mathbf{X}'\mathbf{D}\mathbf{X}).$$

The usual D-optimality criterion value is thus t times larger than the $D_{\boldsymbol{\beta}}$-optimality criterion value in (5). Because t depends on the design through $p_{i,C}$, $p_{i,NC}$ and $V_{i,C}$, the two criteria will, in general, lead to different designs. It can be shown, however, that the two criteria lead to the same design when it is assumed that $\boldsymbol{\beta} = \mathbf{0}_q$ and only designs with $\mathbf{X}'\mathbf{1}_n = \mathbf{0}_q$ are considered.

In the remainder of this paper, we focus on designs that are optimal with respect to the $D_{\boldsymbol{\beta}}$-optimality criterion because, in most applications, the researcher's interest is in $\boldsymbol{\beta}$.

5 Locally optimal designs assuming $\boldsymbol{\beta} = \mathbf{0}_q$

Unlike the information matrix $\mathbf{M}^{0*}$ for the multinomial logit model in (3), the matrix expression in the $D_{\boldsymbol{\beta}}$-optimality criterion for the nested no-choice logit model in (7) is not proportional to the information matrix of a block design in the case of a linear regression model under the assumption that $\boldsymbol{\beta} = \mathbf{0}_q$. However, it does bear strong similarities with it. As a matter of fact, just like for blocked experiments with fixed block effects, it can be shown that the best possible assignment of a given set of alternatives to a given number of choice sets is one that leads to an orthogonally blocked design with blocks of size J (for the conditions for orthogonal blocking, see Goos (2002)). The proof of this result, which is valid for any value of the dissimilarity coefficient λ, is similar to a proof in Goos and Vandebroek (2001) and is given in Appendix B. Even though this result is valid for any type of linear model, we focus on main-effects models in the remainder of the article.

An implication of the proof in Appendix B for main-effects models is that designs for which $\mathbf{X}_i'\mathbf{1}_J = \mathbf{0}_q$ for all choice sets i, i.e. for which every level of a given attribute appears equally often in each choice set, are optimal for a given set of alternatives, $\mathbf{X}$. Furthermore, regular blocked factorial and fractional factorial designs, such as the ones tabulated in Wu and Hamada (2000), are optimal conjoint choice designs among the set of all possible designs if the researcher assumes that $\boldsymbol{\beta} = \mathbf{0}_q$. This is due to the facts that $\det(\mathbf{X}'\mathbf{X})$ is maximal and $\mathbf{X}'\mathbf{1}_n = \mathbf{X}_i'\mathbf{1}_J = \mathbf{0}_q$ for these cases.

The first panel of Table 1 contains an example of a locally D_β-optimal design for a no-choice nested logit model assuming $\boldsymbol{\beta} = \mathbf{0}_6$. The design involves eight choice sets with two real-choice alternatives each and six two-level attributes. The locally optimal design was obtained using the design generators $5 = 123$ and $6 = 124$ and block generators $B_1 = 13$, $B_2 = 23$ and $B_3 = 14$, as suggested in Table 4B.2 in Wu and Hamada (2000), where a list of 16-run fractional factorial two-level designs in blocks is given. A characteristic of the design is that it does not have any attribute level overlap within choice sets, and that the two levels of every attribute occur equally often in every choice set.

6 Bayesian optimal designs

To illustrate the difference between the various approaches for designing conjoint choice experiments with no-choice options, we report some computational results for a choice experiment involving six two-level attributes with eight choice sets containing two real-choice options and one no-choice alternative. The second panel of Table 1 shows a locally D_β-optimal design assuming $\boldsymbol{\beta} = -\mathbf{1}_6$ and $\lambda = 0.5$. The third panel of the table contains a Bayesian D_β-optimal design obtained with a normal prior distribution for $\boldsymbol{\beta}$ with mean $\boldsymbol{\beta}_0 = -\mathbf{1}_6$ and variance $\frac{1}{9}\mathbf{I}_6$, and a uniform prior distribution for λ. The prior distribution utilized for $\boldsymbol{\beta}$ for computing the Bayesian D_β-optimal design basically implies that the signs of the model coefficients are known but not their absolute magnitudes. For searching the Bayesian optimal designs, we followed the approach of Sándor and Wedel (2001, 2005) and Kessels et al. (2006) and sought designs that performed best with respect to the D_β- and D-optimality criteria averaged over 1000 random draws from the prior distribution, and the prior distributions for $\boldsymbol{\beta}$ and λ were assumed independent.

A key difference between the locally and Bayesian D_β-optimal designs in Panels 2 and 3 of Table 1, on the one hand, and the orthogonally blocked design in Panel 1, on the other hand, is that the former designs have a substantial amount of level overlap, whereas the latter design has no attribute level overlap at all. The presence of attribute level overlap is surprising because it is not in line with classical design principles (see, e.g., Kessels et al. (2006, 2008a, 2008b) for similar observations).

To quantify the loss in terms of the D-optimality criterion due to ignoring the no-choice option in the construction of a design, we have also computed a locally D-optimal design as well as a Bayesian D-optimal design for the multinomial logit model. These two designs, which also exhibit attribute level overlap in several choice sets, are displayed in the fourth and fifth panel of Table 1. Regarding the locally D-optimal design in Panel 4, it is important to stress that this is only one of many equivalent locally D-optimal ones. The design displayed is the one that performed best in terms of the D_β-optimality criterion.

It turns out that these designs are substantially less efficient in terms of the D_β-optimality criterion for the no-choice nested logit model than their counterparts in the second and third panel. The relative D_β-efficiency[2] of the Bayesian D_β-optimal design for the no-choice nested logit model (Panel 3) and the Bayesian D-optimal design for the multinomial logit model (Panel 5) is 1.0919 when $\lambda = 0.5$, signifying that taking into account the presence of the no-choice option when designing the experiment leads to a 9.19% gain in efficiency. The D-efficiency of the locally optimal design for the no-choice nested logit model in Panel 2 relative to the one for the multinomial logit model in Panel 4 equals 1.0282. That this relative efficiency is so small is due to the fact that we decided to display the locally D-optimal design that is best in terms of the D_β-optimality criterion. If we had displayed another locally optimal design for the multinomial logit model, the relative efficiency could have been as large as 1.1398. These computational results indicate that it is really important to take into account the no-choice option in the construction of optimal conjoint choice designs when the assumption that $\boldsymbol{\beta} = \mathbf{0}_q$ is no longer made.

7 Information loss due to the no-choice option

Dhar (1997) mentions that a drawback of including a no-choice option in every choice set of the experiment is that it has a negative impact on the information content of the experiment. This is because nothing is learnt regarding the relative attractiveness of the real-choice alternatives whenever a respondent selects the no-choice option. This loss of information can be quantified by comparing the determinant of the information matrix $\mathbf{M}^*$ of the ordinary multinomial logit model in (1) and the D_β-optimality criterion for the no-choice nested logit model in (5). Unfortunately, this comparison, and thus also the information loss due to the no-choice option, depends on all the parameters of the model, which makes it hard to make general statements.

[2]In optimal experimental design, a relative D-efficiency is computed by taking the ratio of the determinants of two competing designs and raising the ratio to the power $1/q$, where q is the number of model parameters of interest.

Table 1: Alternative design options with eight choice sets containing two real-choice options described by six two-level attributes. Panel 1 contains a design that is locally D-optimal for the multinomial logit model and for the no-choice nested logit model when $\boldsymbol{\beta} = \mathbf{0}_6$. Panels 2 and 3 contain designs that are D_β-optimal for the no-choice nested logit model, and Panels 4 and 5 contain designs that are D-optimal for the multinomial logit model.

	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Choice Set	Locally D-Optimal $\boldsymbol{\beta} = \mathbf{0}_6$	Locally D-Optimal $\boldsymbol{\beta} = -\mathbf{1}_6$ $\lambda = 0.5$	Bayesian D-Optimal $\boldsymbol{\beta} \sim N(-\mathbf{1}_6, \frac{1}{9}\mathbf{I_6})$ $\lambda \sim$ Uniform$[0,1]$	Locally D-Optimal $\boldsymbol{\beta} = -\mathbf{1}_6$	Bayesian D-Optimal $\boldsymbol{\beta} \sim N(-\mathbf{1}_6, \frac{1}{9}\mathbf{I_6})$
1	1 1 1 1 1 1	1 2 2 1 2 2	2 2 1 1 1 2	1 1 2 2 2 1	2 1 1 2 2 1
	2 2 2 2 2 2	2 2 1 2 1 2	1 1 2 1 2 2	2 2 1 1 1 2	1 2 2 1 1 2
2	2 1 1 1 2 2	2 1 1 2 2 2	1 2 1 2 2 1	1 1 2 1 2 2	1 2 1 2 1 2
	1 2 2 2 1 1	2 2 2 2 1 1	2 1 2 1 1 2	2 2 1 2 1 1	2 1 2 1 2 1
3	1 2 1 1 2 2	2 2 1 1 1 2	1 2 1 1 2 2	2 1 2 1 2 1	2 1 1 1 2 2
	2 1 2 2 1 1	1 1 2 2 2 1	2 1 2 2 1 1	1 2 1 2 1 2	1 2 2 2 2 1
4	2 2 1 1 1 1	2 2 1 1 2 1	2 1 1 1 2 2	1 2 2 2 2 1	1 2 1 1 2 2
	1 1 2 2 2 2	1 1 2 2 1 2	1 2 2 2 2 1	2 1 1 2 1 2	1 1 2 2 1 1
5	1 1 2 1 2 1	1 2 1 2 2 1	2 1 2 1 2 2	1 2 2 1 2 1	1 1 1 1 2 2
	2 2 1 2 1 2	2 1 2 1 1 2	1 2 1 2 1 2	2 1 1 2 2 2	2 2 1 1 1 1
6	2 1 2 1 1 2	1 2 1 2 1 2	1 2 2 1 1 2	2 1 2 2 1 1	1 1 2 2 2 1
	1 2 1 2 2 1	2 1 2 1 2 1	2 1 1 2 2 1	1 2 1 1 2 2	2 2 1 1 1 2
7	1 2 2 1 1 2	2 1 1 2 2 1	2 2 1 2 2 2	1 1 2 2 1 2	2 1 2 1 1 2
	2 1 1 2 2 1	1 2 2 1 2 2	2 2 2 1 2 1	2 2 1 1 2 1	1 2 1 2 2 1
8	2 2 2 1 2 1	2 1 2 2 2 1	2 2 1 1 2 1	1 1 1 2 2 2	2 1 2 2 1 2
	1 1 1 2 1 2	1 2 2 1 1 2	1 1 2 2 1 2	2 2 2 1 1 1	1 2 2 1 2 1

That the comparison between the two determinants makes sense can be seen by focusing on orthogonally blocked designs under the assumption that $\boldsymbol{\beta} = \mathbf{0}_q$. If we assume that $\boldsymbol{\beta} = \mathbf{0}_q$, then the determinants to compare are the determinants of the matrices in (2) and (6). As for orthogonally blocked designs $\mathbf{X}'\mathbf{1}_n = \mathbf{X}_i'\mathbf{1}_J = \mathbf{0}_q$, the determinant of the information matrix $\mathbf{M}^*$ of the multinomial logit model in (1) equals

$$\left(\frac{1}{J}\right)^q \det(\mathbf{X}'\mathbf{X}),$$

whereas the D_β-optimality criterion for the no-choice nested logit model in (5) equals

$$\left(\frac{J^{\lambda-1}}{J^\lambda+1}\right)^q \det(\mathbf{X}'\mathbf{X}).$$

The ratio of these two values equals

$$\left(\frac{J^\lambda}{J^\lambda+1}\right)^q$$

signifying that the relative D-efficiency of the experiment with the no-choice option in every choice set yields $\{J^\lambda/(J^\lambda+1)\} \times$ 100% of the information contained within an experiment without a no-choice option.

For $\lambda = 1$ (in which case the multinomial logit model without nesting is appropriate), this relative efficiency is $J/(J+1)$ suggesting that a fraction $1/(J+1)$ of the information is lost because of the no-choice option. This means that one third of the information is lost if a no-choice option is added to a choice set of two alternatives if $\lambda = 1$ and $\boldsymbol{\beta} = \mathbf{0}_q$. This result is in line with our intuition because the λ value of one implies that there is no correlation between the real-choice alternatives and the assumption that $\boldsymbol{\beta} = \mathbf{0}_q$ signifies that the utilities of the real-choice options and the no-choice option are equal, so that each of the $J+1$ options in every choice set has the same probability of being selected, namely $1/(J+1)$.

For $\lambda = 0$, the relative efficiency is equal to 1/2 signifying that half the information is lost because of the no-choice option. Also this result is logical, because a zero value for λ and the assumption that $\boldsymbol{\beta} = \mathbf{0}_q$ imply that the real-choice alternatives are indistinguishable. As a result, there are virtually only two choice options in every choice set, namely the no-choice option and a real-choice option. Each of these has a probability of 1/2 of being selected when $\boldsymbol{\beta} = \mathbf{0}_q$ and $\lambda = 0$.

For $\boldsymbol{\beta} = \mathbf{0}_q$ and λ values between 0 and 1, the relative efficiency monotonically increases with λ from 1/2 to $J/(J+1)$, so that the information loss due to the presence of a no-choice option in every choice set lies between $1/(J+1)$ and 1/2. The relative efficiencies are also an increasing function of the number of alternatives in a choice set, J.

Information losses for nonzero values of $\boldsymbol{\beta}$ are harder to compute. For the Bayesian D_β-optimal design in Panel 3 of Table 1, the loss of information is slightly more than 1/3 on average when $\boldsymbol{\beta}$ and λ are randomly and independently drawn from a normal distribution with mean $\boldsymbol{\beta}_0 = -\mathbf{1}_6$ and variance $\frac{1}{9}\mathbf{I}_6$ and from the uniform distribution, respectively. For the locally D_β-optimal design in Panel 2, the information loss decreases from 0.4142 when $\boldsymbol{\beta} = \mathbf{0}_6$ to 0.3554 when $\boldsymbol{\beta} = -\mathbf{1}_6$ for $\lambda = 0.5$. When $\lambda = 0.75$, the information loss lies between 0.3729 (when $\boldsymbol{\beta} = \mathbf{0}_6$) and 0.2957 (when $\boldsymbol{\beta} = -\mathbf{1}_6$). These results show that the information loss does no longer lie between $1/(J+1)$ and 1/2 when $\boldsymbol{\beta} \neq \mathbf{0}_6$, but that cases exist where the loss is smaller than $1/(J+1)$. Thus, the information loss is in certain practical situations smaller than the minimum value one might expect, $1/(J+1)$.

8 Summary

In this article, we derived design criteria for generating optimal conjoint choice designs in the presence of a no-choice option. The criteria were derived under the assumption that a nested logit model was used to analyze data from the choice experiments with a no-choice option. This model is the most appealing of the ones suggested by Haaijer et al. (2001).

We were also able to prove the optimality of orthogonally blocked designs under the classical simplifying assumption that the respondents are indifferent between the attribute level. While this assumption is not very realistic, it is commonly used in the literature on optimal design for conjoint choice experiments and it provides a good starting point for an algorithmic search for optimal designs under more realistic assumptions. We also generated optimal designs for the nested logit model using the computationally intensive Bayesian approach, which has become state-of-the-art in the construction of optimal conjoint choice designs. The Bayesian approach is useful in more realistic scenarios where some prior information about the signs and/or magnitudes of the model coefficients is available. Our results suggest that it is substantially better to take into account the presence of a no-choice option when designing a choice experiment than to ignore it. Thus, when at least some prior information about consumer preferences is available, we strongly recommend to take into account the presence of a no-choice option in every choice set when setting up the choice design. The results in Street and Burgess (2004b), obtained under the simplifying assumption that the respondents are indifferent between the attribute levels, had suggested otherwise.

A final contribution of the paper is that it provides a measure for quantifying the loss of information due to the presence of a no-choice option. The loss of information is an often mentioned drawback of the no-choice option. In the scenario where respondents are indifferent between the real-choice options in the experiments, our measure suggests that the loss of information lies between $1/(J+1)$ and 1/2, where J is the number of real-choice options in a choice set. Because the loss of information depends on the model parameters in general, it is hard to make more general statements however.

The work presented in this paper assumes that the population under study is homogeneous. Obviously, this will be too strong an assumption in many experimental settings. For the points that we wanted to make on the design for experiments with a no-choice option in every choice set, we felt that assuming heterogeneous respondents and using the computationally involved panel mixed logit model would unnecessarily distract the attention from the message we wish to bring.

Appendix A. Information matrix for the no-choice nested logit model

The information matrix on the unknown model parameters $\boldsymbol{\beta}$ and λ for choice set i can be calculated as

$$\mathbf{M}_i = \mathrm{E}\left[\begin{pmatrix} \frac{\partial l_i}{\partial \boldsymbol{\beta}} \\ \frac{\partial l_i}{\partial \lambda} \end{pmatrix}\begin{pmatrix} \frac{\partial l_i}{\partial \boldsymbol{\beta}} \\ \frac{\partial l_i}{\partial \lambda} \end{pmatrix}'\right] = \mathrm{E}\begin{bmatrix} \left(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\right)' & \left(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial l_i}{\partial \lambda}\right)' \\ \left(\frac{\partial l_i}{\partial \lambda}\right)\left(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\right)' & \left(\frac{\partial l_i}{\partial \lambda}\right)\left(\frac{\partial l_i}{\partial \lambda}\right)' \end{bmatrix}.$$

Because

$$\frac{\partial V_{i,C}}{\partial \boldsymbol{\beta}} = \frac{\sum_{j=1}^{J} e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}\mathbf{x}_{ij}}{\sum_{j=1}^{J} e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}} = p_{i,C}^{-1}\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij},$$

$$\begin{aligned}\frac{\partial l_i}{\partial \boldsymbol{\beta}} &= y_{i,C}p_{i,C}^{-1}(\lambda-1)\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij} + \sum_{j=1}^{J} y_{ij}\mathbf{x}_{ij} - \frac{e^{\lambda V_{i,C}}\lambda}{e^{\lambda V_{i,C}}+1}p_{i,C}^{-1}\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij},\\ &= y_{i,C}p_{i,C}^{-1}(\lambda-1)\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij} + \sum_{j=1}^{J} y_{ij}\mathbf{x}_{ij} - \lambda\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij},\end{aligned}$$

and because $y_{i,C}^2 = y_{i,C}$, $y_{i,C}y_{ij} = y_{ij}$ and $(\sum_{j=1}^{J} y_{ij}\mathbf{x}_{ij})(\sum_{j=1}^{J} y_{ij}\mathbf{x}'_{ij}) = \sum_{j=1}^{J} y_{ij}\mathbf{x}_{ij}\mathbf{x}'_{ij}$, we have that

$$\begin{aligned}\left(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\right)' &= y_{i,C}p_{i,C}^{-2}(\lambda-1)^2\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij}\Big)\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}'_{ij}\Big) + p_{i,C}^{-1}(\lambda-1)\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij}\Big)\Big(\sum_{j=1}^{J} y_{ij}\mathbf{x}'_{ij}\Big)\\ &\quad - 2y_{i,C}p_{i,C}^{-1}\lambda(\lambda-1)\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij}\Big)\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}'_{ij}\Big) + p_{i,C}^{-1}(\lambda-1)\Big(\sum_{j=1}^{J} y_{ij}\mathbf{x}_{ij}\Big)\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}'_{ij}\Big)\\ &\quad + \sum_{j=1}^{J} y_{ij}\mathbf{x}_{ij}\mathbf{x}'_{ij} - \lambda\Big(\sum_{j=1}^{J} y_{ij}\mathbf{x}_{ij}\Big)\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}'_{ij}\Big) - \lambda\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij}\Big)\Big(\sum_{j=1}^{J} y_{ij}\mathbf{x}'_{ij}\Big)\\ &\quad + \lambda^2\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij}\Big)\Big(\sum_{j=1}^{J} p_{ij}\mathbf{x}'_{ij}\Big).\end{aligned}$$

As $\mathrm{E}(y_{ij}) = p_{ij}$ and $\mathrm{E}(y_{i,C}) = p_{i,C}$,

$$\begin{aligned}
\mathrm{E}\Big\{\Big(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\Big)\Big(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\Big)'\Big\} &= p_{i,C}^{-1}(\lambda-1)^2\big(\sum_{j=1}^J p_{ij}\mathbf{x}_{ij}\big)\big(\sum_{j=1}^J p_{ij}\mathbf{x}'_{ij}\big) + p_{i,C}^{-1}(\lambda-1)\big(\sum_{j=1}^J p_{ij}\mathbf{x}_{ij}\big)\big(\sum_{j=1}^J p_{ij}\mathbf{x}'_{ij}\big) \\
&- 2\lambda(\lambda-1)\big(\sum_{j=1}^J p_{ij}\mathbf{x}_{ij}\big)\big(\sum_{j=1}^J p_{ij}\mathbf{x}'_{ij}\big) + p_{i,C}^{-1}(\lambda-1)\big(\sum_{j=1}^J p_{ij}\mathbf{x}_{ij}\big)\big(\sum_{j=1}^J p_{ij}\mathbf{x}'_{ij}\big) \\
&+ \sum_{j=1}^J p_{ij}\mathbf{x}_{ij}\mathbf{x}'_{ij} - 2\lambda\big(\sum_{j=1}^J p_{ij}\mathbf{x}_{ij}\big)\big(\sum_{j=1}^J p_{ij}\mathbf{x}'_{ij}\big) + \lambda^2\big(\sum_{j=1}^J p_{ij}\mathbf{x}_{ij}\big)\big(\sum_{j=1}^J p_{ij}\mathbf{x}'_{ij}\big), \\
&= \sum_{j=1}^J p_{ij}\mathbf{x}_{ij}\mathbf{x}'_{ij} + p_{i,C}^{-1}\{\lambda^2(1-p_{i,C})-1\}\big(\sum_{j=1}^J p_{ij}\mathbf{x}_{ij}\big)\big(\sum_{j=1}^J p_{ij}\mathbf{x}'_{ij}\big), \\
&= \mathbf{X}'_i\mathbf{P}_i\mathbf{X}_i + p_{i,C}^{-1}\{\lambda^2 p_{i,NC}-1\}(\mathbf{X}'_i\mathbf{p}_i)(\mathbf{p}'_i\mathbf{X}_i), \\
&= \mathbf{X}'_i\ [\ \mathbf{P}_i + p_{i,C}^{-1}\{\lambda^2 p_{i,NC}-1\}\mathbf{p}_i\mathbf{p}'_i\ ]\ \mathbf{X}_i,
\end{aligned}$$

where $\mathbf{X}_i = [\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iJ}]'$, $\mathbf{P}_i = \mathrm{diag}(p_{i1}, \ldots, p_{iJ})$ and $\mathbf{p}_i = [p_{i1}, \ldots, p_{iJ}]'$. Notice that $\lambda^2 p_{i,NC} - 1 \leq 0$ because $0 \leq p_{i,NC} \leq 1$ and $\lambda \leq 1$.

In addition,

$$\frac{\partial l_i}{\partial \lambda} = y_{i,C}V_{i,C} - \frac{e^{\lambda V_{i,C}}V_{i,C}}{e^{\lambda V_{i,C}}+1} = (y_{i,C}-p_{i,C})V_{i,C},$$

so that

$$\Big(\frac{\partial l_i}{\partial \lambda}\Big)\Big(\frac{\partial l_i}{\partial \lambda}\Big)' = \Big(\frac{\partial l_i}{\partial \lambda}\Big)^2 = (y_{i,C} - 2y_{i,C}p_{i,C} + p_{i,C}^2)V_{i,C}^2$$

and

$$\begin{aligned}
\Big(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\Big)\Big(\frac{\partial l_i}{\partial \lambda}\Big) &= (y_{i,C}p_{i,C}^{-1} - y_{i,C})(\lambda-1)V_{i,C}\sum_{j=1}^J p_{ij}\mathbf{x}_{ij} + (y_{i,C}-p_{i,C})V_{i,C}\sum_{j=1}^J y_{ij}\mathbf{x}_{ij} \\
&- (y_{i,C}-p_{i,C})\lambda V_{i,C}\sum_{j=1}^J p_{ij}\mathbf{x}_{ij}, \\
&= (y_{i,C}p_{i,C}^{-1} - y_{i,C})(\lambda-1)V_{i,C}\sum_{j=1}^J p_{ij}\mathbf{x}_{ij} + (1-p_{i,C})V_{i,C}\sum_{j=1}^J y_{ij}\mathbf{x}_{ij} \\
&- (y_{i,C}-p_{i,C})\lambda V_{i,C}\sum_{j=1}^J p_{ij}\mathbf{x}_{ij},
\end{aligned}$$

where the last step can be made because $y_{i,C}y_{ij} = y_{ij}$. As a result,

$$\mathrm{E}\Big\{\Big(\frac{\partial l_i}{\partial \lambda}\Big)\Big(\frac{\partial l_i}{\partial \lambda}\Big)\Big\} = (p_{i,C} - p_{i,C}^2)V_{i,C}^2 = p_{i,C}p_{i,NC}V_{i,C}^2$$

and

$$\mathrm{E}\Big\{\Big(\frac{\partial l_i}{\partial \boldsymbol{\beta}}\Big)\Big(\frac{\partial l_i}{\partial \lambda}\Big)\Big\} = (1-p_{i,C})(\lambda-1)V_{i,C}\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij} + (1-p_{i,C})V_{i,C}\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij}$$
$$= (1-p_{i,C})\lambda V_{i,C}\sum_{j=1}^{J} p_{ij}\mathbf{x}_{ij},$$
$$= p_{i,NC}\lambda V_{i,C}\mathbf{X}_i'\mathbf{p}_i.$$

The per choice set information matrix is therefore given by

$$\mathbf{M}_i = \begin{bmatrix} \mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i & p_{i,NC}\lambda V_{i,C}\mathbf{X}_i'\mathbf{p}_i \\ p_{i,NC}\lambda V_{i,C}\mathbf{p}_i'\mathbf{X}_i & p_{i,C}p_{i,NC}V_{i,C}^2 \end{bmatrix},$$

where $\mathbf{D}_i = [\ \mathbf{P}_i + p_{i,C}^{-1}\{\lambda^2 p_{i,NC} - 1\}\mathbf{p}_i\mathbf{p}_i'\ ]$.

Appendix B. Optimality of orthogonally blocked designs

Consider a given set of profiles that can be arranged in an orthogonally blocked design with S blocks of size J. Denote the matrix with the attribute levels for a given set of profiles by $\mathbf{X}$ (the SJ rows of which correspond to the profiles) and the J rows of $\mathbf{X}$ assigned to block i by $\mathbf{X}_i$. For any orthogonally blocked arrangement of the profiles, the $D_{\boldsymbol{\beta}}$-optimality criterion under the simplifying assumption that $\boldsymbol{\beta} = \mathbf{0}_q$ can be written as

$$\begin{aligned} D_{\text{orth}} &= \det[\ \mathbf{X}'\mathbf{X} + J^{-1}\{\lambda^2(J^\lambda+1)^{-1}-1\}\sum_{i=1}^{S}\mathbf{X}_i'\mathbf{1}_J\mathbf{1}_J'\mathbf{X}_i \\ &\quad - n^{-1}\lambda^2 J^{\lambda-1}(J^\lambda+1)^{-2}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\ ], \\ &= \det[\ \mathbf{X}'\mathbf{X} + J^{-1}\{\lambda^2(J^\lambda+1)^{-1}-1\}\sum_{i=1}^{S}(S^{-1}\mathbf{X}'\mathbf{1}_n)(S^{-1}\mathbf{1}_n'\mathbf{X}) \\ &\quad - n^{-1}\lambda^2 J^{\lambda-1}(J^\lambda+1)^{-2}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\ ], \\ &= \det[\ \mathbf{X}'\mathbf{X} + n^{-1}S^{-1}\{\lambda^2(J^\lambda+1)^{-1}-1\}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X} \\ &\quad - n^{-1}\lambda^2 J^{\lambda-1}(J^\lambda+1)^{-2}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\ ], \\ &= \det[\ \mathbf{X}'\mathbf{X} + g\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\ ], \end{aligned} \tag{8}$$

where

$$g = n^{-1}S^{-1}\{\lambda^2(J^\lambda+1)^{-1}-1\} - n^{-1}\lambda^2 J^{\lambda-1}(J^\lambda+1)^{-2}.$$

This is due to the fact that

$$\mathbf{X}_i'\mathbf{1}_J = S^{-1}\mathbf{X}'\mathbf{1}_n, \quad i = 1, \ldots, S,$$

for an orthogonally blocked design (see, for example, Goos (2002)). Note that g is always negative.

For an arrangement of the profiles in $\mathbf{X}$ that is not orthogonally blocked, we have that

$$\mathbf{X}_i'\mathbf{1}_J = S^{-1}\mathbf{X}'\mathbf{1}_n + \boldsymbol{\delta}_i, \quad i = 1, \ldots, S,$$

with

$$\sum_{i=1}^{S} \boldsymbol{\delta}_i = \mathbf{0}_q.$$

In that case, the $D_{\boldsymbol{\beta}}$-optimality criterion under the simplifying assumption that $\boldsymbol{\beta} = \mathbf{0}_q$ can be written as

$$\begin{aligned}
D_{\text{n.orth}} &= \det[\, \mathbf{X}'\mathbf{X} + J^{-1}\{\lambda^2(J^\lambda+1)^{-1} - 1\} \sum_{i=1}^{S} (S^{-1}\mathbf{X}'\mathbf{1}_n + \boldsymbol{\delta}_i)(S^{-1}\mathbf{1}_n'\mathbf{X} + \boldsymbol{\delta}_i') \\
&\quad - n^{-1}\lambda^2 J^{\lambda-1}(J^\lambda+1)^{-2}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\,], \\
&= \det[\, \mathbf{X}'\mathbf{X} + J^{-1}\{\lambda^2(J^\lambda+1)^{-1} - 1\}\{\sum_{i=1}^{S} S^{-2}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X} + \sum_{i=1}^{S} \boldsymbol{\delta}_i\boldsymbol{\delta}_i' \\
&\quad + \sum_{i=1}^{S} S^{-1}\mathbf{X}'\mathbf{1}_n\boldsymbol{\delta}_i' + \sum_{i=1}^{S} S^{-1}\boldsymbol{\delta}_i\mathbf{1}_n'\mathbf{X}\} - n^{-1}\lambda^2 J^{\lambda-1}(J^\lambda+1)^{-2}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\,], \qquad (9) \\
&= \det[\, \mathbf{X}'\mathbf{X} + g\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X} \\
&\quad + J^{-1}\{\lambda^2(J^\lambda+1)^{-1} - 1\}\{\sum_{i=1}^{S} \boldsymbol{\delta}_i\boldsymbol{\delta}_i' + S^{-1}\mathbf{X}'\mathbf{1}_n \sum_{i=1}^{S} \boldsymbol{\delta}_i' + S^{-1}(\sum_{i=1}^{S} \boldsymbol{\delta}_i)\mathbf{1}_n'\mathbf{X}\}\,], \\
&= \det[\, \mathbf{X}'\mathbf{X} + g\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X} + J^{-1}\{\lambda^2(J^\lambda+1)^{-1} - 1\} \sum_{i=1}^{S} \boldsymbol{\delta}_i\boldsymbol{\delta}_i'\,].
\end{aligned}$$

Now, because $\sum_{i=1}^{S} \boldsymbol{\delta}_i\boldsymbol{\delta}_i'$ is nonnegative definite and because $J^{-1}\{\lambda^2(J^\lambda+1)^{-1} - 1\}$ is always negative, the determinant in (9) will always be smaller than the one in (8). This proves that orthogonal arrangements of given designs $\mathbf{X}$ ought to be preferred if it is assumed that $\boldsymbol{\beta} = \mathbf{0}_q$.

If the given design $\mathbf{X}$ maximizes $\det(\, \mathbf{X}'\mathbf{X} + g\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\,)$ and it can be arranged so that it forms an orthogonally blocked design with S blocks of size J, then this arrangement is an optimal conjoint choice design under the assumption that $\boldsymbol{\beta} = \mathbf{0}_q$.

References

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*, Oxford U.K.: Clarendon Press.

Burgess, L. and Street, D. J. (2003). Optimal designs for 2^k choice experiments, *Communications in Statistics: Theory and Methods* **32**: 2185–2206.

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

Dhar, R. (1997). Consumer preference for a no-choice option, *Journal of Consumer Research* **24**: 215–231.

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

Goos, P. and Vandebroek, M. (2001). D-optimal response surface designs in the presence of random block effects, *Computational Statistics and Data Analysis* **37**: 433–453.

Grasshoff, U., Grossmann, H., Holling, H. and Schwabe, R. (2003). Optimal paired comparison designs for first-order interactions, *Statistics* **37**: 373–386.

Grasshoff, U., Grossmann, H., Holling, H. and Schwabe, R. (2004). Optimal designs for main effects in linear paired comparison models, *Journal of Statistical Planning and Inference* **126**: 361–376.

Grasshoff, U., Grossmann, H., Holling, H. and Schwabe, R. (2007). Design optimality in multi-factor generalized linear models in the presence of an unrestricted quantitative factor, *Journal of Statistical Planning and Inference* **137**: 3882–3893.

Grossmann, H., Holling, H., Grasshoff, U. and Schwabe, R. (2006). Optimal designs for asymmetric linear paired comparisons with a profile strength constraint, *Metrika* **64**: 109–119.

Grossmann, H., Holling, H. and Schwabe, R. (2002). Advances in optimum experimental design for conjoint analysis and discrete choice models, *in* P. H. Franses and A. L. Montgomery (eds), *Advances in Econometrics 16: Econometric Models in Marketing*, Amsterdam: JAI, pp. 93–117.

Haaijer, R., Kamakura, W. and Wedel, M. (2001). The no-choice alternative in conjoint choice experiments, *International Journal of Market Research* **43**: 93–106.

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.

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

Kessels, R., Jones, B., Goos, P. and Vandebroek, M. (2008a). An efficient algorithm for constructing Bayesian optimal choice designs. To appear in *Journal of Business and Economic Statistics*.

Kessels, R., Jones, B., Goos, P. and Vandebroek, M. (2008b). Recommendations on the use of Bayesian optimal designs for choice experiments, *Quality and Reliability Engineering International* **24**: 737–744.

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

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

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

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

Street, D. J. and Burgess, L. (2004a). Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments, *Journal of Statistical Planning and Inference* **118**: 185–199.

Street, D. J. and Burgess, L. (2004b). Optimal stated preference choice experiments when all choice sets contain a specific option, *Statistical Methodology* **1**: 37–45.

Train, K. E. (2003). *Discrete Choice Models with Simulation*, Cambridge: Cambridge University Press.

van Berkum, E. E. M. (1987). Optimal paired comparison designs for factorial and quadratic models, *Journal of Statistical Planning and Inference* **15**: 265–278.

van Berkum, E. E. M. and Pauwels, B. (2003). D-optimal weighted paired comparison designs, *Journal of Statistical Planning and Inference* **110**: 147–157.

Wu, C. F. J. and Hamada, M. (2000). *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: Wiley.