

Construction of Heterogeneous Conjoint Choice Designs: A New Approach

Qing Liu

Wisconsin School of Business, University of Wisconsin–Madison, Madison, Wisconsin 53706, qliu@bus.wisc.edu

Yihui (Elina) Tang

College of Business Administration, University of Illinois at Chicago, Chicago, Illinois 60607, etang@uic.edu

Extant research on choice designs in marketing focuses on the construction of efficient homogeneous designs where all respondents get the same design. Recently marketing scholars proposed the construction of efficient heterogeneous designs where different respondents or groups of respondents get different subdesigns, and demonstrated substantial efficiency gain when such heterogeneous designs are employed. A significant hurdle in the widespread adoption of heterogeneous designs is the high computation cost, even when the number of subdesigns contained in the heterogeneous design is restricted to be small. In this paper we propose a new approach for the construction of efficient heterogeneous choice designs. In contrast to extant approaches that are based on an exact design framework where it is computationally prohibitive to do an exhaustive search to find a globally optimal design, our proposed approach is based on the continuous design framework where well-established mathematical theories can be leveraged for quick identification of a globally optimal design. The proposed approach makes it feasible to generate a highly efficient choice design that is completely heterogeneous—a unique subdesign for each individual respondent in the choice experiment. The proposed approach is the first in the marketing literature to find a completely heterogeneous choice design with assured high global design efficiency using the continuous design framework. Results from simulation and empirical studies demonstrate superior performance of the proposed approach over extant approaches in constructing efficient heterogeneous choice designs.

Data, as supplemental material, are available at <http://dx.doi.org/10.1287/mksc.2014.0897>.

Keywords: experimental design; conjoint analysis; conjoint choice designs; discrete choice designs; optimal designs; mixed logit model

History: Received: June 17, 2013; accepted: October 21, 2014; Preyas Desai served as the editor-in-chief and Harald van Heerde served as associate editor for this article. Published online in *Articles in Advance* March 19, 2015.

1. Introduction

Conjoint choice experiments have been widely used by practitioners and academics in marketing research to measure consumer preferences. In a conjoint choice experiment, respondents are given multiple sets of product alternatives and are asked to pick their most preferred product from each choice set. These product alternatives are constructed using various combinations of product attributes and their levels. The responses obtained from the conjoint choice experiment are then used to estimate respondents' preferences for product attributes and their levels. Based on this information, marketers can identify characteristics of products that affect consumers' choice/buying behavior and use it for a variety of applications that include new product/concept evaluation, repositioning, competitive analysis, pricing, and market segmentation (see review by Green and Srinivasan 1990).

An important issue to consider when designing conjoint choice experiments is the statistical efficiency of the design. That is, how to construct the multiple

sets of product alternatives so that respondents' preferences can be estimated efficiently. The D-criterion has been used in assessing the efficiency of an experimental design on parameter estimation and has been the main criterion used in constructing efficient choice designs. Formally, it is defined as the minimization of the determinant of the inverse of the Fisher information matrix, also called the D-error in the marketing literature (e.g., Huber and Zwerina 1996; Arora and Huber 2001; Sándor and Wedel 2002, 2005), and aims to minimize the volume of the confidence region for the parameter estimates. Alternative design criteria have also been proposed in the marketing literature, including the A-criterion (Kuhfeld et al. 1994), which aims to minimize the average variance of the parameter estimates, the M-criteria (Toubia and Hauser 2007), which aim at efficient estimation of linear functions of model parameters that are managerially relevant, and the V- and G-criteria (Kessels et al. 2006), which aim at accurate prediction of future choices rather than efficient estimation of model parameters that capture consumer preferences on product attributes and

attribute levels. Another stream of research focuses on increasing the response quality in preference measurement (see discussions by Netzer et al. 2008) by better engaging the respondents through incentive-aligned mechanisms (Ding et al. 2005, Ding 2007) or game-like mechanisms (Ding et al. 2009, Toubia et al. 2012).

Extant research on choice designs in marketing concentrates on constructing a homogeneous design—all respondents get the same design, and thus everyone gets the same sets of product alternatives for evaluation. Sándor and Wedel (2005) were the first to propose the use of heterogeneous designs where different respondents or groups of respondents get different designs (or subdesigns) in the choice experiment. They obtained heterogeneous designs that contain two to 10 subdesigns using a computer search algorithm that relied on the swapping and cycling of attribute levels. Even though the number of subdesigns was small in the heterogeneous designs they constructed, Sándor and Wedel (2005) demonstrated that there is substantial efficiency gain where a heterogeneous design is used instead of a homogeneous design.

A significant hurdle in widespread adoption of heterogeneous designs in practice is the high computation cost—the search for an efficient heterogeneous design involves the search of not one, but multiple (M) subdesigns. The size of a heterogeneous design is therefore much larger than a homogeneous design (i.e., M times the size of a homogeneous design). Computationally, this may become cost prohibitive. To reduce the computation time, Sándor and Wedel (2005) recommended the sequential (or greedy) search for each of the subdesigns or the separate search of all subdesigns independently. Even with these shortcuts, the total computation cost for the search of an efficient heterogeneous design is still at least M times the computation cost for the search of an efficient homogeneous design. Although many existing computer search algorithms for constructing efficient choice designs work well when constructing homogeneous designs, they are inadequate for the case of heterogeneous designs. The computation time required to search for an efficient heterogeneous design is simply too long, especially when there is a large number of subdesigns contained in the heterogeneous design. As a result, marketing practitioners often resort to randomized designs by Sawtooth software, where the subdesigns are generated in a balanced but somewhat randomized way that aims for one-way and two-way balances over the frequency of occurrences of different attribute levels while allowing for some repeated occurrences of the same attribute level within a choice task (see Chrzan and Orme 2000). Such designs “work” but their design feasibility comes at the expense of design efficiency.

We propose a new, flexible, and cost-effective approach specific for the construction of efficient heterogeneous choice designs. Based on the continuous design framework (see, for example, Silvey 1980, Pukelsheim 1993) and recent research on optimal designs in the statistics literature (Yang et al. 2013), the proposed approach makes it computationally feasible to obtain heterogeneous designs with assured high design efficiency. We show through examples that compared to efficient heterogeneous choice designs obtained from the separate search approach recommended by Sándor and Wedel (2005), designs obtained through our proposed approach achieve efficiency gains from 12.5% to 16.6% (when measured by the D-error), and at the same time take only a fraction (12% to 20%) of the computation time. In contrast to the separate search approach that restricts the number of subdesigns to be small (e.g., six), the proposed approach does not impose such restrictions and generates a completely heterogeneous design with a unique subdesign for each individual respondent. Although the Sawtooth randomized designs widely used in practice are also completely heterogeneous, their design efficiency is shown to be significantly lower than that of the heterogeneous designs obtained from our proposed approach.

Methodologically, our proposed approach makes a significant contribution because it is based on the continuous design framework, unlike extant approaches to efficient choice designs that are based on the exact design framework. In contrast to the exact design framework where it is computationally prohibitive to do an exhaustive search over all possible designs to find a design that is globally optimal, the continuous design framework has well-established mathematical theories and tools to leverage for the fast identification of a globally optimal continuous design. In particular, we discuss in §3 the use of the general equivalence theorem (Kiefer 1974) and our extension of the optimal weight exchange algorithm by Yang et al. (2013) to achieve this goal.

Our proposed approach makes it feasible to generate a completely heterogeneous design with assured high global design efficiency. In contrast to extant approaches to efficient heterogeneous designs that explicitly search for a given number of subdesigns (e.g., six), our proposed approach generates heterogeneous designs without imposing any restrictions on the number of subdesigns. In other words, whereas designs obtained from these extant approaches are only heterogeneous at the group level where different groups of respondents get different subdesigns, our proposed approach makes it practical to generate a completely heterogeneous design where each respondent is given a unique subdesign without additional computation cost. To the best of our knowledge, our

proposed approach is the first in the marketing literature to find a completely heterogeneous choice design with assured high global design efficiency through the use of the continuous design framework.

Our research is both distinctly different from and complementary to previous research on adaptive choice designs (e.g., Toubia et al. 2004, 2007). Unlike adaptive choice designs where the choice sets are sequentially constructed and customized on the fly for each respondent based on the respondent's responses to previous questions, our research focuses on designs where the choice sets are constructed all at once before the choice experiment is launched and responses collected, as commonly used in practice.

The remainder of the paper is organized as follows. In §2 we provide an overview of extant approaches to efficient choice designs. We then provide a detailed description of our new approach in §3 and report results from the comparative study in §4. We present an empirical application of the proposed approach in §5, and discuss potential limitations and extensions of the proposed approach in §6. We end the paper with a summary and conclusion in §7.

2. Review on Extant Approaches to Efficient Choice Designs

According to McFadden's (1974) random utility model of consumer choice, the utility u_{sj} for alternative j in choice set s ($j = 1, \dots, J$ and $s = 1, \dots, S$) is

$$u_{sj} = x'_{sj}\beta + \varepsilon_{sj}, \quad (1)$$

where x_{sj} is a vector that captures the characteristics of alternative j in choice set s , and correspondingly β is the vector of parameters (part-worths). Note that when a no-choice option is used in the conjoint choice experiment for each choice set, it is treated as a separate alternative with zero utility by setting the vector $x_{sj} = 0$, and correspondingly the number of part-worths β is increased by one (Haaijer et al. 2001, Gilbride and Allenby 2006). The errors $\{\varepsilon_{sj}\}$ are assumed to have type 1 extreme value distribution with location parameter 0 and scale parameter 1.

2.1. Efficient Choice Designs for the Mixed Logit Model

Recognizing the importance of accounting for heterogeneity in modeling consumer behavior in marketing, recent research on efficient choice designs focuses on the mixed logit model (e.g., Sándor and Wedel 2002, 2005; Yu et al. 2009; Liu and Arora 2011). In contrast to the standard logit model that assumes β is homogeneous across consumers, the mixed logit model accounts for consumer heterogeneity by assuming that

the m parameters in β are random effects. For example, for the choice designs investigated in Sándor and Wedel (2002, 2005), it is assumed that $\beta = \mu_B + V\sigma_B$ with vector $\sigma_B = (\sigma_1, \dots, \sigma_m)'$, and V is an $m \times m$ diagonal matrix, where the m diagonal elements are independent and identically distributed (i.i.d.) from the standard normal distribution. The probability that alternative j is chosen from choice set s , given μ_B and σ_B , is

$$\varphi_{sj} = \int p_{sj}(v)f(v)dv, \quad \text{where}$$

$$p_{sj}(v) = \frac{\exp\{x'_{sj}(\mu_B + V\sigma_B)\}}{\sum_{j=1}^J \exp\{x'_{sj}(\mu_B + V\sigma_B)\}}, \quad (2)$$

where v is the vector containing the m diagonal elements of matrix V .

Under the mixed logit model, an optimal design for efficient estimation of the parameters $\theta = (\mu_B, \sigma_B)$ under the D-criterion is the one that minimizes the D-error (Sándor and Wedel 2002) defined as the determinant of the inverse of the Fisher information matrix normalized by the total number of elements in θ (which is $2m$ in this case), that is

$$D\text{-error} = \det\{I_\theta(X_1, \dots, X_H)^{-1}\}^{1/(2m)}, \quad \text{with}$$

$$I_\theta(X_1, \dots, X_H) = \sum_{h=1}^H \sum_{s=1}^S I_\theta(X_{hs}), \quad (3)$$

where X_h is the design matrix corresponding to the S choice sets of product alternatives administered to respondent h for evaluation, and $I_\theta(X_{hs})$ is the Fisher information matrix for a single choice set s ($s = 1, 2, \dots, S$) for respondent h ($h = 1, 2, \dots, H$). In the case of homogeneous designs, $X_1 = X_2 = \dots = X_H$, the Fisher information matrix in (3) simplifies to

$$I_\theta(X_1, \dots, X_H) = H \sum_{s=1}^S I_\theta(X_s)$$

$$= H \sum_{s=1}^S \begin{bmatrix} E'_s \Delta_s^{-1} E_s & E'_s \Delta_s^{-1} Q_s \\ Q'_s \Delta_s^{-1} E_s & Q'_s \Delta_s^{-1} Q_s \end{bmatrix}, \quad (4)$$

$$E_s = \int [P_s(v) - p_s(v)p_s(v)'] X_s f(v) dv,$$

$$Q_s = \int [P_s(v) - p_s(v)p_s(v)'] X_s V f(v) dv,$$

$$p_s(v) = (p_{s1}(v), \dots, p_{sj}(v))',$$

$$P_s(v) = \text{diag}(p_{s1}(v), \dots, p_{sj}(v)),$$

$$\Delta_s = \text{diag}(\varphi_{s1}, \dots, \varphi_{sj}).$$

Note that the expression of the Fisher information matrix for the choice model, as shown in (4), involves the vector of choice probabilities, which are

a function of the model parameters. Therefore, choice designs cannot be optimized without a prior estimate of the model parameters (Ben-Akiva and Lerman 1985, Chapter 8.5). Drawing from insights obtained from earlier research on choice designs (Huber and Zwerina 1996, Arora and Huber 2001), one solution to the problem is to use prior knowledge to specify the values of $\theta = (\mu_B, \sigma_B)$ for the design construction. Alternatively, instead of using the fixed values or point-mass prior estimates of $\theta = (\mu_B, \sigma_B)$, the Bayesian approach can be used to incorporate the uncertainty of the prior estimates through the prior probability distributions and correspondingly the D-criterion becomes the Bayesian D-criterion (see Chaloner and Verdinelli 1995 for a review).

2.2. Computer Search Algorithms

Various computer search algorithms have been used in extant research to find efficient choice designs, such as the RS (relabeling and swapping) algorithm (Huber and Zwerina 1996), the RSC (relabeling, swapping, and cycling) algorithm (Sándor and Wedel 2002), the modified Fedorov algorithm (Kessels et al. 2006), and the coordinated exchange algorithm (Kessels et al. 2009, Yu et al. 2009, Liu and Arora 2011). The most recent research has advocated the use of the coordinate-exchange algorithm because it is found to be computationally efficient, especially when there is a large number of attributes and attribute levels in the choice experiment. All of these algorithms start with an initial choice design and then go through various exchanges of attribute levels (i.e., an attribute level in the initial choice design is exchanged with all other possible levels of that attribute). An exchange is accepted only if it improves the design criterion value. The exchanges are done iteratively until no further substantial improvement is possible.

Because the choice design under search in these algorithms is discrete in nature, there is no mathematical tool or theory we can leverage to ensure that the final design obtained is globally optimal. That is, it is difficult to claim that no other design is better, unless we check every possible option in the entire design space. However, the number of all possible choice designs can be so large that it is impossible to search over every possible design and find a globally optimal design. For example, for a choice experiment with four attributes each with three levels, there are $3^4 = 81$ possible product alternatives or attribute-level combinations to choose from. Suppose we want to construct a choice design with three alternatives per choice set, then the number of all possible choice sets is

$$\binom{\text{number of all possible product alternatives}}{\text{number of alternatives per choice set}} = \binom{81}{3} = 85,320. \quad (5)$$

If we try to construct a homogeneous design with eight choice sets, then the number of all possible choice designs is

$$\binom{\text{number of all possible choice sets}}{\text{number of choice sets in the choice design}} = \binom{85,320}{8} = 6.9 \times 10^{34}. \quad (6)$$

The number of all possible choice designs will be even larger if we try to construct a heterogeneous design. With the rare exception of an extremely simple case, it is impossible for the computer search to be exhaustive and search over all possible choice designs to find a globally optimal design. Thus, a common problem with the currently available searching algorithms mentioned earlier is that the search may give rise to a design that is far less efficient than a globally optimal design. To remedy this problem, multiple searches (tries) with different initial designs have been used in practice and the best design is selected (e.g., Kessels et al. 2009, Yu et al. 2009, Liu and Arora 2011). Although design efficiency is improved, with such an approach the computation time can be very long for the search of homogeneous designs. The computation time would be undoubtedly even longer for the search of heterogeneous designs.

Moreover, there are two outstanding questions with this approach of multiple searches: First, how many different initial designs are sufficient? Should we include 50, 100, 1,000, or more tries in the design search? When can we stop and say that that is enough? There is no definite answer from extant research—different researchers have used different numbers of tries in their search of efficient designs. For example, Kessels et al. (2009) used 250 tries, Yu et al. (2009) used 1,000 tries, and Liu and Arora (2011) used 100 tries. Second, and perhaps more importantly, how can we be assured that the final design obtained is highly efficient? All we can claim is that the obtained design has higher efficiency than the initial designs and all other designs covered in the computer search during the exchange process. However, no matter how many different initial designs we try, there is no guarantee that we can find a globally optimal design unless all possible designs are covered in the computer search. Besides, without the knowledge of a globally optimal design, the true or global efficiency (i.e., efficiency relative to a globally optimal design) of the design obtained from the computer search is unknown.

3. Proposed Approach

We propose a novel approach for the construction of heterogeneous choice designs that effectively addresses the limitations of the extant approaches

outlined above. The proposed approach makes it computationally feasible to find a completely heterogeneous choice design with assured high global efficiency. For ease of illustration, we focus on the D-criterion in this section, and discuss in §6 how the approach can be easily extended to other design criteria such as the A-, V-, and M-criteria.

3.1. The Continuous Design Framework

The proposed approach employs the continuous design (also called the approximate design) framework (see Atkinson et al. 2007, Chapter 9), which is different from the exact design framework used in extant approaches in the marketing literature. Unlike an exact design, which depends on the specific number of observations in the choice experiment, a continuous design is based on the probability measure (or weights) associated with the choice sets and is independent of the total number of observations. In particular, a continuous design in the conjoint choice context can be denoted as $\{(C_k, w_k)\}$, $k = 1, \dots, K$, where C_k is a choice set that belongs to the space of all possible choice sets, and w_k is the corresponding weight for the choice set with the constraint such that $0 \leq w_k \leq 1$ and $\sum_{k=1}^K w_k = 1$. Note that the weight w_k is continuous and thereby the corresponding design is defined in the continuous space rather than the exact space. The Fisher information matrix for a continuous design $\{(C_k, w_k)\}$, $k = 1, \dots, K$ is defined as $\sum_{k=1}^K w_k I(C_k)$, where $I(C_k)$ is the Fisher information matrix for the single choice set C_k that contains J product alternatives, that is, $C_k = (x_{k1}, x_{k2}, \dots, x_{kJ})'$. To distinguish a continuous design from an exact design, we use the notation \tilde{X} to represent a continuous design in contrast to the corresponding exact design X for the remainder of the paper.

For any exact design $X = \{C_{hs}\}$, $h = 1, \dots, H$, $s = 1, \dots, S$, where C_{hs} denotes the choice set s for respondent h , there exists a corresponding continuous design $\tilde{X} = \{(C_{hs}, 1/(HS))\}$, $h = 1, \dots, H$, $s = 1, \dots, S$. This is because there are a total of HS observations, and therefore the weight corresponding to each choice set C_{hs} is $1/HS$. Given a fixed number of respondents and a fixed number of choice sets per respondent, the Fisher information matrix of an exact design X , which equals to $\sum_{h=1}^H \sum_{s=1}^S I(C_{hs})$, is proportional to that of its corresponding continuous design \tilde{X} , which equals to $\sum_{h=1}^H \sum_{s=1}^S (1/(HS))I(C_{hs})$. Let G represent the globally optimal exact design. The global efficiency of an exact design X , defined as the efficiency relative to the optimal exact design G , is therefore the same as the relative efficiency of the two designs' continuous counterparts \tilde{X} and \tilde{G} . For example, under the D-criterion for the mixed logit model, let $D\text{-error}(X)$ and $D\text{-error}(G)$ denote the D-errors (as defined in Equations (3) and (4)) associated with the exact design

X and the globally optimal exact design G , and let $D\text{-error}(\tilde{X})$ and $D\text{-error}(\tilde{G})$ denote the D-errors associated with the corresponding continuous designs, we have

Global efficiency of an exact design X

$$= \frac{D\text{-error}(G)}{D\text{-error}(X)} = \frac{D\text{-error}(\tilde{G})}{D\text{-error}(\tilde{X})}. \quad (7)$$

Note that \tilde{G} is the continuous counterpart of the globally optimal exact design G , and it may or may not be globally optimal in the continuous design space. Let \tilde{X}^* be the globally optimal continuous design such that $D\text{-error}(\tilde{X}^*)$ is minimized over the entire space of continuous designs, then we have $D\text{-error}(\tilde{G}) \geq D\text{-error}(\tilde{X}^*)$, and therefore from (7) we have

Global efficiency of an exact design X

$$\geq \frac{D\text{-error}(\tilde{X}^*)}{D\text{-error}(\tilde{X})}. \quad (8)$$

The right-hand side of (8) measures the global efficiency of design \tilde{X} in the continuous design space, and it provides a lower bound to the global efficiency of the corresponding design X in the exact design space, as shown by the above inequality. The advantage of using the right-hand side of (8) to assess global design efficiency in contrast to (7) is that it does not require the knowledge of a globally optimal exact design G , which is computationally infeasible to obtain, as discussed in §2. Instead, we can focus on finding a globally optimal continuous design \tilde{X}^* , where the well-established general equivalence theorem by Kiefer (1974) can be used to help achieve this goal.

3.2. The General Equivalence Theorem

The general equivalence theorem provides the necessary and sufficient condition for a globally optimal continuous design. We can verify the global optimality of any design by verifying this condition. Note that there may be multiple designs that are globally optimal as long as they all satisfy this condition. These designs are "equivalent" in the sense that they all have the same value of the selected optimality criterion (e.g., they all have the same minimum D-error values). The theorem is a general theorem that applies to a wide class of design optimality criteria, including the D-, A-, V-, and M-design criteria based on the Fisher information matrix. Here we provide a brief review of the theorem for the D-criterion in the context of choice designs.

General Equivalence Theorem. A design \tilde{X}^* is globally optimal in the continuous design space if, and only if,

$$d(C_k, \tilde{X}^*) \leq 0, \quad (9)$$

for any C_k that belongs to the space of all possible choice sets. For the D-criterion, we have

$$d(C_k, \tilde{X}^*) = \text{Tr}\{[I(C_k) - I(\tilde{X}^*)]I(\tilde{X}^*)^{-1}\}, \quad (10)$$

where $I(C_k)$ is the Fisher information matrix for choice set C_k , and $I(\tilde{X}^*)$ is the Fisher information matrix for the continuous design \tilde{X}^* .

For a design criterion other than the D-criterion, the expression of the function $d(C_k, \tilde{X}^*)$ in the general equivalence theorem differs. We provide in Appendix A the expressions of the function $d(C_k, \tilde{X}^*)$ for the A-, V-, and M-design criteria. Detailed proofs of the theorem can be found in Kiefer (1974), Pukelsheim (1993), Fedorov and Hackl (1997), and Atkinson et al. (2007), among others. The significance of the general equivalence theorem is that it allows us to check whether a design is indeed globally optimal by verifying if the necessary and sufficient condition (9) is satisfied over the space of all possible choice sets. That is, to find a globally optimal continuous design using the general equivalence theorem, we only need to check over the space of all possible choice sets. In contrast, to find a globally optimal exact design, we need to check over the space of all possible choice designs. The number of all possible choice sets is obviously much smaller than that of all possible choice designs. For example, in the design scenario discussed in §2.2, there are four attributes each with three levels, three alternatives per choice set, and eight choice sets in the design. The number of all possible choice sets is 85,320 (see Equation (5)), a much smaller number in comparison to the number of all possible choice designs, which is 6.9×10^{34} (see Equation (6)). Therefore, in contrast to the computationally prohibitive task of finding a globally optimal exact design, it is much easier to find a globally optimal continuous design. In fact, the computation time is very fast using the algorithm, which we describe in §3.3.

3.3. Algorithm for Finding a Globally Optimal Continuous Design

Recently, following the line of work by Yang and Stufken (2009) who obtained a series of unifying results for multiple models and multiple design optimality criteria, Yang et al. (2013) proposed a new optimal weight exchange algorithm (OWEA) for deriving globally optimal designs in the continuous design space. Instead of relying completely on numeric computation, the OWEA utilizes Newton's iteration method, a well-established mathematical tool, to find

optimal designs for a broad class of optimality criteria. The authors theoretically proved the convergence of the algorithm and demonstrated that the proposed algorithm substantially outperforms existing algorithms in terms of computation time.

The problem considered by Yang et al. (2013) focuses on nonlinear models with a small number of continuous variables in which they can utilize the finer grid approach as an adaptive way to simplify the search process. In this paper we extend the OWEA to the setting of choice experiments with a large number of discrete variables (various attribute levels) where the finer grid approach in OWEA cannot be applied and we develop a new approach to address this issue. We call our algorithm the modified OWEA (mOWEA). The algorithm works by iteratively updating the choice sets and their corresponding weights until they converge to a globally optimal continuous design such that the sufficient and necessary condition (9) in the general equivalence theorem is satisfied. Figure 1 outlines the four steps of the mOWEA, which can be summarized as follows:

(1) Start by randomly selecting r choice sets¹ from all possible choice sets and assign equal weights for each selected choice set, where r is the total number of parameters in the model. This serves as the initial design.

(2) For a given initial design \tilde{X}_0 , update the initial weights corresponding to the selected choice sets with the optimal weights using Newton's iteration method (details of the method are provided in Appendix B). Note that there may be multiple iterations with Newton's iteration method before the optimal weights are found. Elimination of zero, one, or multiple choice sets from the initial design may occur during the process.

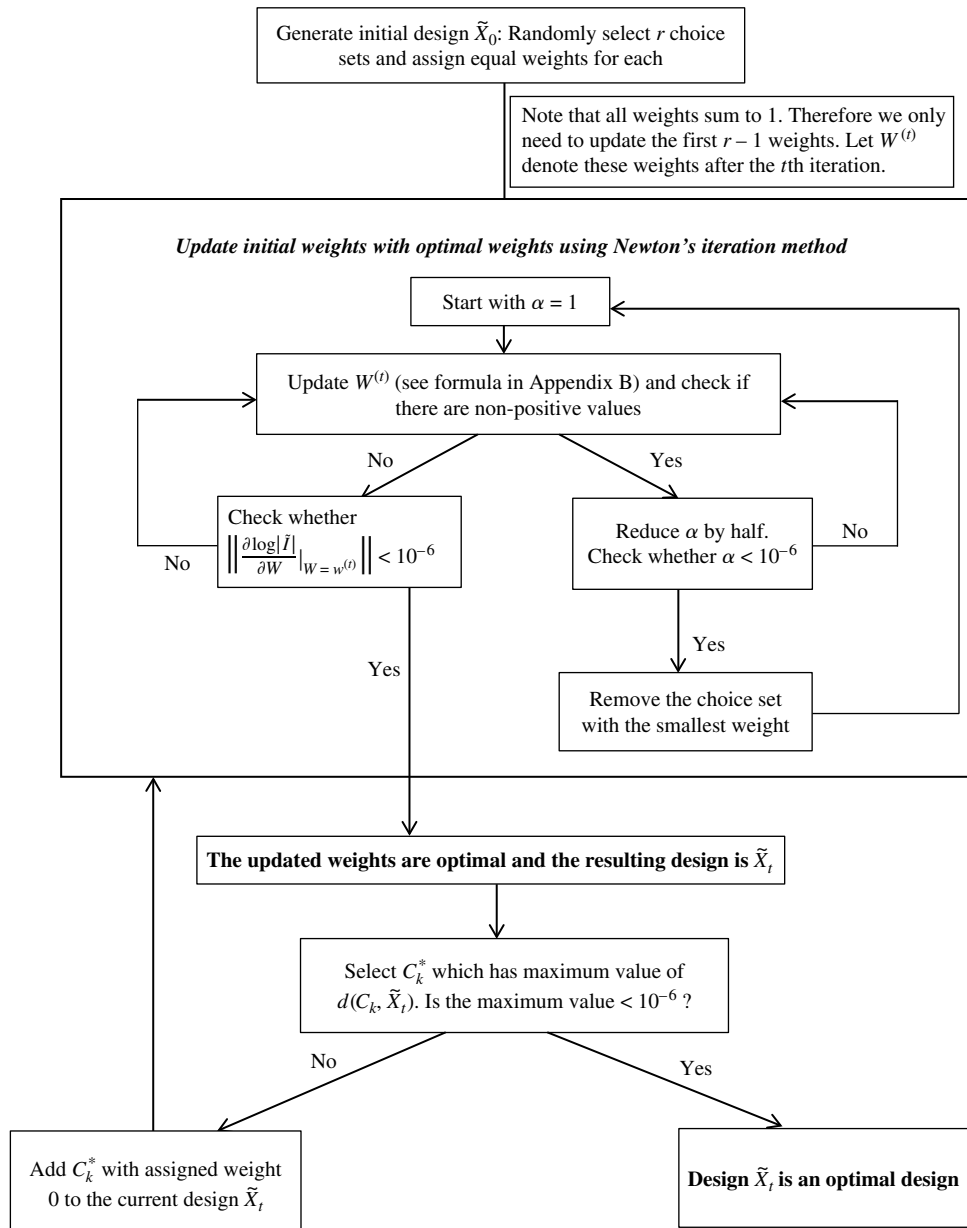
(3) For the updated design \tilde{X}_t , calculate the function $d(C_k, \tilde{X}_t)$ as defined in Equation (10) for every possible choice set C_k and find the choice set C_k^* which maximizes $d(C_k, \tilde{X}_t)$.

If the maximum value is less than a prespecified threshold (e.g., 10^{-6}), we consider the necessary and sufficient condition in Equation (9) satisfied and therefore \tilde{X}_t is a globally optimal continuous design.

(4) Otherwise, update the design \tilde{X}_t by adding to the design the choice set C_k^* with assigned weight 0. This serves as the new initial design, and repeat steps (2) and (3).

¹In theory, the number of choice sets in the initial design can be set to any number. Regardless of which initial design to use, the final converged optimal continuous design should be equivalent. However, we advise against starting with an initial design where the number of choice sets is large (e.g., 1,000) because this will significantly slow down the computation speed in the next step because of the calculation of the large Hessian matrix in Newton's iteration method (see details in Appendix B).

Figure 1 Flow Chart of the Algorithm



To illustrate the use of the algorithm through an example, consider constructing a globally optimal continuous design that involves four product attributes each with three levels and three product alternatives per choice set. The design optimization criterion used is the D-criterion for the mixed logit model as discussed in §2.1. The total number of parameters for the mixed logit model is $8(\mu_B) + 8(\sigma_B) = 16$. Thus, the algorithm starts by randomly selecting 16 choice sets from the 85,320 possible choice sets (see Equation (5)), and allocating an equal weight (1/16) to each choice set. These initial weights are updated to the optimal weights according to the Newton iteration method. Then for the updated

design \tilde{X}_t , calculate the function $d(C_k, \tilde{X}_t)$ for all possible choice sets ($k = 1, \dots, 85,320$). Add to the design \tilde{X}_t the choice set C_k^* that maximizes $d(C_k, \tilde{X}_t)$ with an initial weight of 0. This becomes the new initial design. Iteratively update the weights and the choice sets until the necessary and sufficient condition in Equation (9) is satisfied and the resulting design is globally optimal.

3.4. Conversion from Optimal Continuous Design to Exact Design with High Efficiency

Note that the globally optimal continuous design obtained from the algorithm described above is independent of the number of respondents and the

Downloaded from informs.org by [128.248.155.225] on 09 April 2016, at 19:45. For personal use only, all rights reserved.

number of choice sets per respondent. In this section we describe a procedure we developed to convert the optimal continuous design to an exact design with assured high efficiency. With the weight w_k being continuous between 0 and 1, the continuous design framework does not limit the number of observations per design point to be integers. That is, when the total number of observations is N , where $N = \text{number of respondents } (H) \times \text{number of choice sets per respondents } (S)$, the number of observations $Nw_k = HS w_k$ required for choice set C_k is not necessarily an integer. Therefore, in order to convert the continuous design to an exact design, the number of observations required for each choice set needs to be rounded to an integer. Although there are existing rounding approaches in the statistics literature (see, for example, Pukelsheim 1993, p. 311) that convert an optimal continuous design to an exact design with high efficiency, these approaches do not work well in the setting of choice designs because they do not impose any restrictions on the replication of the design points. As a result, a respondent may be asked to evaluate two or more choice sets that are exactly the same, which clearly does not make sense. Therefore, we develop a procedure that addresses this issue, with the assumption that the maximum weight in the globally optimal continuous design is less than or equal to $1/S$ (which we have found to hold true for all design scenarios explored in our simulation studies described in §4). The steps of the procedure are outlined below.

1. Rank order (from high to low) the choice sets within the globally optimal continuous design by their corresponding required number of observations $\tilde{l}_{0k} = HS w_k$, which are not necessarily integers. Assign the first S choice sets to the first respondent ($h = 1$).

2. For the S choice sets that have been assigned in the previous step, let $\tilde{l}_{hk} = \tilde{l}_{h-1,k} - 1$. For those that have not been assigned, let $\tilde{l}_{hk} = \tilde{l}_{h-1,k}$. Reorder the choice sets according to \tilde{l}_{hk} and assign the first S choice sets to the next respondent (i.e., respondent $h + 1$).

3. Repeat step 2 until we have assigned S choice sets for each of the H respondents.

To see the basic intuition of the procedure, consider an example of converting the globally optimal continuous design obtained in §3.3 to an exact design for the case of 120 respondents and eight choice sets per respondent. The required number of observations for choice set C_k is $\tilde{l}_{0k} = HS w_k = 960 w_k$, according to the globally optimal continuous design. The converting procedure starts by assigning the top eight choice sets with the highest \tilde{l}_{0k} to the first respondent. Then the number of observations to be allocated to each choice set is adjusted to reflect the fact that these eight choice sets have already been assigned once (such that $\tilde{l}_{1k} = \tilde{l}_{0k} - 1$) although others have

not (such that $\tilde{l}_{1k} = \tilde{l}_{0k}$). The choice sets are subsequently reordered according to the adjusted number of observations \tilde{l}_{1k} and the top eight choice sets are assigned to the next respondent. This continues until each of the 120 respondents is assigned with eight choice sets. The procedure ensures the following two nice properties of the resulting exact design: (i) no respondent will be given duplicate choice sets, and (ii) the resulting exact design is assured to have high global efficiency. It is easy to see that the first property holds because by definition of the procedure, no choice set can be taken more than once for the same respondent. The second property is ensured because the corresponding continuous design of the resulting exact design is very close to the globally optimal continuous design. In other words, the weight of each choice set of the continuous design corresponding to the resulting exact design is very close to the optimal weight w_k in the globally optimal continuous design. We refer readers to Appendix C for the detailed proof.

Note that when a no-choice option is used in the conjoint choice experiment, we need to add the no-choice option as an additional alternative to each choice set of the resulting exact design from the above procedure. In addition, in calculating the Fisher information matrix, the no-choice option is treated as a separate alternative with zero utility ($x_{sj} = 0$) together with an increased number of parameters for estimation (Haaijer et al. 2001, Gilbride and Allenby 2006), e.g., an increase from $8(\mu_B) + 8(\sigma_B) = 16$ to $9(\mu_B) + 9(\sigma_B) = 18$ parameters. All other aspects of the proposed approach remain the same.

We end this section by noting that in addition to the fast computation and the assured high global efficiency of the resulting choice design, our proposed approach naturally gives rise to completely heterogeneous designs. In particular, whereas the existing approach to efficient heterogeneous designs (Sándor and Wedel 2005) explicitly searches for a small number of (e.g., six) subdesigns to be administered to different groups of respondents and thus results in designs that are only heterogeneous at the group level, our approach generates completely heterogeneous designs that are not limited to a prespecified number of subdesigns.

4. Comparative Study: Simulations

In this section, we compare the proposed approach with extant approaches in the search of efficient heterogeneous choice designs through a series of simulation studies. We focus on the D-criterion for the mixed logit model by minimizing the D-error as expressed in (3) and (4), and discuss extensions to other design criteria later in §6. We start by providing an overview of extant approaches to the construction of heterogeneous choice designs.

4.1. Extant Approaches to Heterogeneous Choice Designs

As discussed in Sándor and Wedel (2005), extant computer-search approaches to efficient heterogeneous choice designs include (i) the simultaneous search, (ii) the greedy (sequential) search, and (iii) the separate search. These three approaches rank from the most to the least computationally intensive. Specifically, for a heterogeneous design with M subdesigns each with S choice sets, the simultaneous search refers to the search of a big design with $M \times S$ choice sets, the greedy search refers to the sequential search of each of the M subdesigns, and the separate search refers to the independent search of the M subdesigns. Because of the high computation cost, it is recommended by Sándor and Wedel (2005) to limit the number of subdesigns to six, with the insights that marginal efficiency gain becomes small once the number of subdesigns exceeds six. Among the three approaches, the separate search is found to have the highest computational efficiency, while at the same time produces designs with similar efficiency as designs from the other two approaches. Therefore, the separate search is the recommended approach for finding heterogeneous designs for the mixed logit model by Sándor and Wedel (2005) and we use it as the main benchmark in our comparative study.

In regards to the specific computer-search algorithm for the separate search approach, we use the coordinate exchange algorithm (Meyer and Nachtsheim 1995), which has been proven to be fast and efficient based on recent research on the construction of efficient choice designs (Kessels et al. 2009, Yu et al. 2009, Liu and Arora 2011). The coordinate exchange algorithm works by exchanging each attribute level in the initial design with all possible levels of that attribute and accepting the exchange only if it improves the design criterion value. Starting with the first attribute of the first alternative in the design, the algorithm iterates until no further substantial improvement is possible. Without a clear guideline on the number of tries (where each try refers to a search with one different initial design) to be used for the coordinate exchange algorithm, we take the two extremes used in extant literature, that is, 100 tries on one end as used by Liu and Arora (2011) and 1,000 tries on the other end as used by Yu et al. (2009), and examine the impact of the different number of tries on both the computation time of the search and the efficiency of the resulting design.

We also include in our comparison the Sawtooth randomized designs as the industry benchmark for heterogeneous choice designs. In contrast to the proposed approach and the separate search approach that aim to find designs optimized under the D-criterion

for the mixed logit model, the Sawtooth randomized designs are constructed in a balanced but somewhat randomized way that aims for one-way and two-way balances over the frequency of occurrences of different attribute levels while allowing for some repeated occurrences of the same attribute level within a choice task (i.e., “balanced overlap;”² see Chrzan and Orme 2000).

4.2. Performance Comparison Results

For ease of comparison, we start with the same two design settings as those investigated by Sándor and Wedel (2005), that is, four attributes each with three levels, three alternatives per choice set, and the number of choice sets S for each respondent is either $S = 10$ (for design setting 1) or $S = 15$ (for design setting 2). Including the two settings allows for investigation on whether the number of choice sets has any effect on the efficiency comparisons between the proposed approach and the benchmark approaches. Without loss of generality, we set the number of respondents H in the choice experiment to 120. Both the Sawtooth randomized designs and the designs obtained from our proposed approach contain $M = 120$ subdesigns each to be allocated to one individual respondent. In contrast, the separate search approach generates six subdesigns where each subdesign is to be allocated to a group of 20 respondents.

Note that both the separate search approach from Sándor and Wedel (2005) and our proposed approach require prior knowledge of the model parameters for the design construction. We follow Sándor and Wedel (2002, 2005) and assume that $\mu_B = (-1, 0, -1, 0, -1, 0)$, and $\sigma_B = (1, 1, \dots, 1)$. In addition, we use the same 64-point quasi-random sample for both approaches in the evaluation of integrals in the Fisher information matrix, as shown in (4). The 64-point quasi-random sample is obtained through the Halton sequence approach, which has been shown by Yu et al. (2009) to be more efficient than alternative random or quasi-random samples. Using the same sample ensures that the integrals in the Fisher information matrix are evaluated with the same accuracy across both approaches under comparison, and thus any differences we may observe in computation time or design efficiency are indeed attributable to the nature of the different approaches rather than different evaluations of the integrals.

²The “balanced overlap” is recommended by Sawtooth software as the preferred method to use for the generation of randomized designs. The other main method used by Sawtooth software for this purpose is the “complete enumeration” method, which allows for no repeated occurrence of the same attribute level (i.e., no attribute-level overlap) within a choice task.

4.2.1. Performance Comparison on Global Design Efficiency and Relative Efficiency. We start by comparing the proposed approach with the benchmark approaches based on the following two measures of design efficiency:

(1) *Global design efficiency.* This is a measure of absolute design efficiency, as defined on the right-hand side of Equation (8). In particular, it assesses the efficiency of a given design over the entire space of continuous designs by calculating the ratio of the D-error of the given design and that of the globally optimal continuous design.

(2) *Percentage improvement in efficiency.*³ This is a measure of relative design efficiency that has been used in the marketing literature (e.g., Sándor and Wedel 2001, 2002, 2005; Yu et al. 2009) to assess the performance of a given design (X) relative to a reference design (or baseline design, R). It is calculated as

$$\text{Percentage improvement in efficiency} = \left[1 - \frac{\text{D-error}(X)}{\text{D-error}(R)} \right] \times 100. \quad (11)$$

If the measure is positive, then it shows that design X is more efficient than the reference design. Otherwise it is less efficient. For our comparative study, we use the design obtained from the separate search with 100 tries as our reference design and calculate percentage improvement in efficiency for all other designs relative to this design.

The top panel of Table 1 shows the comparison results based on global design efficiency. The heterogeneous designs obtained from the proposed approach have almost 100% global design efficiency (99.998% and 99.995%, respectively), confirming the assured high efficiency of the designs obtained from the proposed approach. In contrast, designs obtained from benchmark approaches have much lower efficiency. Specifically, the separate search approach results in heterogeneous designs with global design

³ This measure has also been called “percentage reduction in sample size” in previous research that focuses on homogeneous designs. This is because the Fisher information in the definition of the D-error is proportional to the sample size given that the design is fixed and independent of the sample size, such as in the case of a homogeneous design where every respondent gets the same design. Thus, if a design is twice as efficient as the reference design based on the D-error, then we can reduce the number of respondents by half when using this design and still obtain parameter estimates that are as efficient as those from the reference design. However, in the case of a heterogeneous design, this does not apply because the design is changing with the number of respondents, that is, in contrast to the case of a homogeneous design where every respondent gets the same design and thus reducing the sample size does not change the design, reducing the sample size (e.g., from 200 to 100 respondents) in the case of a heterogeneous design changes the design itself because each respondent gets a different subdesign.

efficiency ranging from 83.37% to 88.79%, and the Sawtooth randomized designs are at 55.43% and 55.28% efficiency,⁴ respectively, for the two design settings. Results on the relative design efficiency measure reconfirm the high efficiency of the heterogeneous designs from the proposed approach. As shown in the middle panel of Table 1, designs obtained from the proposed approach are 12.51% and 16.63% more efficient in comparison to the designs obtained from the separate search approach with 100 tries. On the other hand, the Sawtooth randomized designs are 50.41% and 58.27% less efficient.

It should be noted that although the proposed approach and the separate search approach both aim to find designs specifically optimized under the D-criterion for the mixed logit model, the Sawtooth randomized designs aim for balanced overlap (Chrzan and Orme 2000) instead. Therefore the lower efficiency of the Sawtooth randomized designs should not come as a surprise. Further, the proposed approach and the separate search approach assume $\mu_B = (-1, 0, -1, 0, -1, 0, -1, 0)$ and $\sigma_B = (1, 1, \dots, 1)$ in the construction of optimal designs for the mixed logit model. In contrast, the Sawtooth randomized designs do not have the capability to incorporate such prior knowledge and instead assume by default that all respondents have zero part-worths with no heterogeneity such that all product alternatives have equal utilities and equal probabilities of being chosen (Sawtooth software technical support). As a result, Sawtooth designs can be highly efficient for settings of zero part-worths with no or minimal heterogeneity, but their design efficiency diminishes in the presence of high heterogeneity and nonzero part-worths as seen in our comparative study. Our result is consistent with findings in Yu et al. (2009, Table 4) as well, where the Sawtooth design (the “nearly orthogonal design” in the table) was found to have 35% (= 0.288/0.818) efficiency relative to the design optimized for the mixed logit model that accounts for consumer heterogeneity and nonzero part-worths (the “locally D-optimal mixed logit design”).

4.2.2. Performance Comparison on Computation Time. In addition to comparisons based on global design efficiency and relative design efficiency, we also make comparisons between the proposed

⁴ We also examined the Sawtooth randomized designs generated by the “complete enumeration” method and found that they are at 39.51% and 39.56% global design efficiency, respectively, for the two design settings. As shown by Sándor and Wedel (2002), efficient choice designs for the mixed logit model require some attribute-level overlap within the choice sets. The complete enumeration method allows for no overlap, and therefore results in randomized designs with lower efficiency in comparison to randomized designs generated by the balanced overlap method.

Table 1 Comparison Between the Proposed Approach and Benchmark Approaches

	Number of choice sets	Proposed approach ($M = 120$)	Separate search approach ($M = 6$)		Sawtooth randomized design ($M = 120$)
			100 tries	1,000 tries	
Global design efficiency	$S = 10$	99.998%	83.37%	87.03%	55.43%
	$S = 15$	99.995%	87.49%	88.79%	55.28%
Percentage improvement in efficiency	$S = 10$	16.63%	—	4.21%	–50.41%
	$S = 15$	12.51%	—	1.46%	–58.27%
Computation time (hh:mm:ss)	$S = 10$	00:02:10	00:10:43	01:46:51	00:00:04
	$S = 15$	00:02:15	00:17:31	04:04:15	00:00:05

Notes. (i) Both design settings (“ $S = 10$ ” and “ $S = 15$ ”) involve four attributes each with three levels; three alternatives per choice set; 120 respondents. (ii) The computation time for Sawtooth randomized designs is not directly comparable to other approaches because of the different computing platforms used—whereas the Sawtooth randomized designs are obtained exclusively from the Sawtooth software, designs from the proposed approach and the separate search approach are obtained using SAS/IML on a Dell computer with 2.2 GHz and 8 GB RAM.

approach and benchmark approaches based on computation time. The bottom panel of Table 1 shows the time taken, in the format of hours:minutes:seconds (hh:mm:ss), to obtain heterogeneous designs for each design setting. We note that as an industry benchmark, the randomized designs by Sawtooth software are very fast to generate (four and five seconds for the two design settings). However, the computation time is not directly comparable to other approaches under comparison because of the different computing platform used. In particular, while the Sawtooth randomized designs are obtained exclusively from Sawtooth software, designs from the proposed approach and the separate search approach are obtained using SAS/IML on a Dell computer with 2.2 GHz and 8 GB RAM. Therefore, we next focus on the comparison between the proposed approach and the separate search approach where the computation time is directly comparable.

We find that there is a substantial gain in computational efficiency using the proposed approach in comparison to the separate search approach. As shown in the bottom panel of Table 1, the time taken by the proposed approach is approximately only 20% and 12% of that taken by the separate search approach with 100 tries for the settings of 10 choice sets and 15 choice sets, respectively. Note that the number of subdesigns is limited to six for the separate search approach, whereas it is not limited for the proposed approach. This suggests that the gain in computational efficiency of the proposed approach would be considerably larger if the limit on the number of subdesigns is relaxed for the separate search approach.

Another advantage of the proposed approach over the benchmark separate search approach is that there is a minimal change in computation time when the number of choice sets per respondent increases. As shown in the bottom panel of Table 1, there is an increase of only five seconds in computation time for the proposed approach when the number of choice sets per respondent increases from 10 to

15. In contrast, the computation time for the separate search approach with 100 tries has an increase of nearly seven minutes. This is because the proposed approach is based on the continuous design framework where the search of an optimal continuous design is independent of the number of choice sets and respondents. It is only during the conversion from the optimal continuous design to the exact design that the number of choice sets makes a difference together with the number of respondents. The computation time required for the conversion procedure is minimal and therefore an increased number of choice sets per respondent has little influence on the computation time for the proposed approach.

Last, we note that in the separate search approach, the computation time increases considerably as the number of tries increases from 100 to 1,000. This, in combination with the minimal differences in design efficiency between designs obtained from 100 tries and 1,000 tries (as shown in the top panel of Table 1), strongly argues for the use of 100 tries rather than 1,000 tries. Therefore, we use 100 tries only for the separate search approach in the remainder of our comparative study.

4.3. Comparison Results Under Different Parameter Settings

The comparison results discussed so far are obtained under a standard setting of parameters μ_B and σ_B in the mixed logit model used in previous research (e.g., Sándor and Wedel 2002, 2005), that is, $\mu_B = (-1, 0, -1, 0, -1, 0, -1, 0)$, and $\sigma_B = (1, 1, \dots, 1)$. Next we extend the comparative study to different parameter settings and examine if and how the performance of the proposed approach changes accordingly. Given that the construction of Sawtooth randomized designs is based on the assumption of zero part-worths, we start by examining two settings in which the mean part-worths are zero ($\mu_B = (0, 0, \dots, 0)$), with either standard consumer heterogeneity ($\sigma_B = (1, 1, \dots, 1)$) or in the

Table 2(a) Efficiency Results (in %) Under Zero Mean Part-Worths

Design		Zero mean part-worths $\mu_B = (0, \dots, 0)$	
		Standard heterogeneity $\sigma_B = (1, \dots, 1)$	Minimal heterogeneity $\sigma_B = (0.01, \dots, 0.01)$
Global design efficiency	Proposed approach	99.99	99.99
	Separate approach	87.18	96.27
	Sawtooth randomized	58.80	82.92
Percentage improvement in efficiency	Proposed approach	12.81	3.72
	Separate approach	—	—
	Sawtooth randomized	-48.27	-16.10

Note. Results reported in the table are for the design setting “ $S = 15$ ” that involves four attributes each with three levels; three alternatives per choice set; 15 choice sets per respondent; 120 respondents.

Table 2(b) Efficiency Results (in %) Under Various Nonzero Part-Worths and Heterogeneity

Design		Low heterogeneity ($\lambda = 0.5$)		High heterogeneity ($\lambda = 2$)	
		Low magnitude ($\delta = 0.5$)	High magnitude ($\delta = 2$)	Low magnitude ($\delta = 0.5$)	High magnitude ($\delta = 2$)
		Global design efficiency	Proposed approach	99.99	99.99
Separate approach	91.87		86.93	82.98	80.29
Sawtooth randomized	68.29		49.55	45.11	39.67
Percentage improvement in efficiency	Proposed approach	8.13	13.07	17.02	19.71
	Separate approach	—	—	—	—
	Sawtooth randomized	-34.52	-75.44	-83.95	-102.39

extreme case with nearly zero heterogeneity⁵ ($\sigma_B = (0.01, 0.01, \dots, 0.01)$). This allows us to investigate whether the benefit of the proposed approach over the Sawtooth randomized designs still holds and how it is affected by the zero prior specification.

In addition, we examine various settings of nonzero part-worths and investigate how the performance of the proposed approach changes when levels of consumer heterogeneity and/or magnitudes of consumer part-worths increase from low to high. Specifically, we examine $2 \times 2 = 4$ different settings. We have low and high levels of consumer heterogeneity, corresponding to $\lambda = 0.5$ and 2, respectively, where the vector of consumer heterogeneity parameters is $\sigma_B = \lambda \times (1, 1, \dots, 1)$. We also have low and high magnitudes of mean part-worths, corresponding to $\delta = 0.5$ and 2, respectively, where the vector of mean part-worths in the mixed logit model is $\mu_B = \delta \times (-1, 0, -1, 0, -1, 0, -1, 0)$.

For each setting, we obtain heterogeneous designs with 15 choice sets per respondent using the proposed approach and benchmark approaches for comparison. The computation time taken to generate the designs are similar to those reported in Table 1. We report results on global and relative design efficiency measures in Tables 2(a) and 2(b). The results

in both tables show that designs obtained from the proposed approach consistently outperform the benchmark designs obtained from the separate search approach and the Sawtooth randomized approach. For example, designs from the proposed approach achieve 99.99% global design efficiency across all settings. In contrast, the global design efficiency ranges from 80.29% to 96.27% for designs from the separate search approach, and 39.67% to 82.92% for Sawtooth randomized designs.

In comparison to the results reported in Table 1, results in Table 2(a) show that under standard consumer heterogeneity, the efficiency gain of the proposed approach over the Sawtooth randomized approach is not affected much by the change from the nonzero prior to the zero prior specification of the mean part-worths. This is because there is only a slight change (from 55.28% to 58.80%) in global design efficiency for the Sawtooth randomized design. In the extreme case when there is almost no consumer heterogeneity, the global design efficiency improves for the Sawtooth randomized design. Nevertheless, the efficiency gain from using the proposed approach is still about 17% (99.99% over 82.92%) based on the D-error.

The results in Table 2(b) demonstrate that in general, the efficiency gain from using the proposed approach increases when either consumer heterogeneity or magnitude of part-worths increases. The magnitude of the increase in efficiency gain from using

⁵ The heterogeneity parameters in σ_B in the mixed logit model cannot be set to exactly 0 because this will make the information matrix singular.

the proposed approach, however, is much higher corresponding to the increase in consumer heterogeneity (e.g., an increase from 8.13% to 17.02%, in contrast to an increase from 8.13% to 13.07%). This suggests that the advantage of using the proposed approach in finding efficient heterogeneous choice designs would be most prominent when consumer heterogeneity is high.

4.4. Comparison Results Under Alternative Measures of Efficiency

The comparison results in the previous two subsections provide strong supporting evidence in favor of the proposed approach over benchmark approaches in finding efficient heterogeneous choice designs for the mixed logit model. In this section, we further examine the performance of the designs under comparison based on alternative measures of efficiency. We start with the following two measures that have often been used in extant research (e.g., Sándor and Wedel 2002, Kessels et al. 2006, Yu et al. 2009).

(a) *Efficiency on prediction.* This alternative measure focuses on predictive accuracy and is based on the expected root mean squared prediction error (ERMSE_p). In particular, suppose there are Q all possible future choice sets each with J alternatives, the ERMSE_p for a given design X is defined as

$$\text{ERMSE}_p(X) = \frac{1}{Q \times J} \int \{[\varphi(\hat{\theta}) - \varphi(\theta)]' [\varphi(\hat{\theta}) - \varphi(\theta)]\}^{1/2} \cdot f(\hat{\theta}) d\hat{\theta}, \quad (12)$$

where $f(\hat{\theta})$ is the distribution of the parameter estimates, $\varphi(\hat{\theta})$ is the vector of predicted probabilities for the mixed logit model computed according to Equation (2) using the parameter estimates $\hat{\theta} = (\hat{\mu}_B, \hat{\sigma}_B)$, and $\varphi(\theta)$ is the vector of true probabilities computed using the true parameter values. The smaller the ERMSE_p, the more efficient the corresponding design is in predictive accuracy. Using the design obtained from the separate search approach as the reference design R , we assess the relative efficiency of a given design X in predictive accuracy by calculating the efficiency gain relative to the reference design as follows:

$$\begin{aligned} &\text{Efficiency gain in prediction} \\ &= \left[1 - \frac{\text{ERMSE}_p(X)}{\text{ERMSE}_p(R)} \right] \times 100. \quad (13) \end{aligned}$$

A positive value on the efficiency gain suggests that design X is more efficient than the reference design in predictive accuracy, whereas a negative value suggests the opposite.

(b) *Efficiency on recovery of the parameters $\theta = (\mu_B, \sigma_B)$ in the mixed logit model.* In contrast to the two design efficiency measures in §4.2.1 that are based on

the D-error, which focuses on the confidence region of the parameter estimates, this alternative measure is based on the ERMSE pertaining to the parameters $\theta = (\mu_B, \sigma_B)$. Specifically, it is defined as

$$\text{ERMSE}_\theta(X) = \int [(\hat{\theta} - \theta)'(\hat{\theta} - \theta)]^{1/2} f(\hat{\theta}) d\hat{\theta}. \quad (14)$$

The smaller the ERMSE_θ, the more efficient the corresponding design is in parameter recovery. Similarly, using the design obtained from the separate search approach as the reference design R , we assess the relative efficiency of a given design X in parameter recovery through the efficiency gain defined the same way as in (13) by replacing the ERMSE_p's with the ERMSE_θ's of the corresponding designs.

Following Sándor and Wedel (2002, 2005) and Yu et al. (2009), we approximate the ERMSEs in Equations (12) and (14) by averaging over a large number of random draws (e.g., 1,000 draws) from the asymptotic distribution of the parameter estimates, i.e., the multivariate normal distribution with the mean vector $\theta = (\mu_B, \sigma_B)$ and the variance-covariance matrix that is equal to the inverse of the Fisher information matrix (Chaloner and Verdinelli 1995). To test if the efficiency gains based on the ERMSEs are statistically significant, we repeat the evaluation multiple times (e.g., 100 times) with a different set of random draws each time.

Although the focus of this paper is on the mixed logit model where the parameters of interest are the population-level parameters $\theta = (\mu_B, \sigma_B)$ that capture mean part-worths and standard deviations, individual-level part-worth estimates can also be obtained through hierarchical Bayes estimation once response data are available. Therefore, it would also be interesting to examine how efficient the designs are in the recovery of individual-level part-worths. To do this, we follow Arora and Huber (2001) to first simulate responses $\mathbf{y} = (y_1, \dots, y_H)$ for all H individual respondents and then use hierarchical Bayes estimation to obtain 25,000 posterior draws of individual-level part-worth estimates $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_H)$. Based on the posterior draws, we calculate the ERMSE pertaining to the individual-level part-worths $\boldsymbol{\beta}$ as

$$\text{ERMSE}_\beta(X) = \int [(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]^{1/2} f(\hat{\boldsymbol{\beta}} | \mathbf{y}) d\hat{\boldsymbol{\beta}}. \quad (15)$$

The efficiency gain in recovery of individual-level part-worths is then calculated in the same way as in (13) by replacing the ERMSE_p's with the ERMSE_β's of the corresponding designs.

Tables 3(a) and 3(b) show the performance comparisons based on these alternative measures of efficiency for various settings of mean part-worths and consumer heterogeneity. On both the predictive accuracy and the recovery of $\theta = (\mu_B, \sigma_B)$, we find that

Table 3(a) Efficiency Results (in %) Under Alternative Measures of Efficiency

Design	Nonzero mean part-worths	Zero mean part-worths	Zero mean part-worths	
	Standard heterogeneity	Standard heterogeneity	Minimal heterogeneity	
Efficiency gain in prediction	Proposed approach	3.07 ^a	9.84 ^a	0.79 ^a
	Separate approach	—	—	—
	Sawtooth randomized	-52.67 ^a	-62.32 ^a	-2.96 ^a
Efficiency gain in recovery of $\theta = (\mu_B, \sigma_B)$	Proposed approach	5.05 ^a	9.10 ^a	4.82 ^a
	Separate approach	—	—	—
	Sawtooth randomized	-45.12 ^a	-39.35 ^a	-23.07 ^a
Efficiency gain in recovery of individual-level part-worths	Proposed approach	1.13	-5.08	-0.41
	Separate approach	—	—	—
	Sawtooth randomized	-3.35	-0.86	-4.29

^aSignificantly different from 0 (α -level = 0.05).

Table 3(b) Efficiency Results (in %) Under Alternative Measures of Efficiency—Various Nonzero Mean Part-Worths and Heterogeneity

Design	Low heterogeneity ($\lambda = 0.5$)		High heterogeneity ($\lambda = 2$)		
	Low magnitude ($\delta = 0.5$)	High magnitude ($\delta = 2$)	Low magnitude ($\delta = 0.5$)	High magnitude ($\delta = 2$)	
	Efficiency gain in prediction	Proposed approach	5.26 ^a	0.11	8.06 ^a
Separate approach		—	—	—	—
Sawtooth randomized		-44.01 ^a	-48.71 ^a	-47.54 ^a	-57.64 ^a
Efficiency gain in recovery of $\theta = (\mu_B, \sigma_B)$	Proposed approach	6.35 ^a	5.21 ^a	9.83 ^a	11.32 ^a
	Separate approach	—	—	—	—
	Sawtooth randomized	-32.73 ^a	-45.23 ^a	-50.94 ^a	-61.83 ^a
Efficiency gain in recovery of individual-level part-worths	Proposed approach	5.42	-2.28	2.30	5.20
	Separate approach	—	—	—	—
	Sawtooth randomized	-3.43	-24.67 ^a	-0.31	2.90

^aSignificantly different from 0 (α -level = 0.05).

the heterogeneous design obtained from the proposed approach significantly outperforms the benchmark designs, especially when there is high consumer heterogeneity. For example, as shown in Table 3(a), for the setting of nonzero mean part-worths and standard heterogeneity, the design from the proposed approach achieves 3.07% efficiency gain in prediction and 5.05% efficiency gain in the recovery of $\theta = (\mu_B, \sigma_B)$ relative to the design obtained from the separate search approach. In contrast, the Sawtooth randomized design results in efficiency losses of 52.67% and 45.12%, respectively. We note that the magnitude of efficiency gains of the proposed design over benchmark designs under these alternative measures of efficiency is smaller than those based on the D-error in Table 1 (e.g., 5.05% versus 12.51%). This is expected because the proposed design is optimized based on the D-error and therefore may not necessarily achieve maximal efficiency gains when evaluated under alternative measures (Kessels et al. 2006).

On the recovery of the individual-level part-worths, we find that there are generally no significant differences among the three types of designs except for one occasion. This should not come as a surprise because the designs from the proposed approach and the

separate search approach are optimized for the estimation of population-level parameters in the mixed logit model, which may not be optimal for the estimation of individual-level parameters (Liu et al. 2012). Therefore, these designs may not outperform the Sawtooth randomized design on the recovery of the individual-level part-worths. However, as shown in Table 3(b), in the occasion when there are high magnitudes of mean part-worths and low levels of heterogeneity, the designs from the proposed approach and the separate search approach do significantly outperform the Sawtooth randomized design. This makes intuitive sense because the Sawtooth design does not take into account the nonzero prior, whereas the other two types of designs do, and the greatest benefit of incorporating the nonzero prior is achieved in the occasion with high magnitudes of mean part-worths and low heterogeneity (Arora and Huber 2001).

In summary, the findings from the comparative study make a strong case for the use of the proposed approach in the construction of efficient heterogeneous choice designs for the mixed logit model. In comparison to the separate search approach that explicitly searches for a small number of sub-designs (e.g., six), our proposed approach makes it

computationally feasible to obtain a completely heterogeneous design with assured high efficiency. In comparison to Sawtooth randomized designs, which serve as the industry standard, heterogeneous designs obtained from our proposed approach achieve considerable gains in design efficiency, especially when consumer heterogeneity is high.

5. Empirical Application of the Proposed Approach

A key takeaway from the comparative study in §4 is that the proposed approach offers a practical and valuable alternative to marketing practitioners in the construction of heterogeneous choice designs. Instead of using Sawtooth randomized designs, which have been the industry standard, marketing practitioners can possibly employ the proposed approach to obtain heterogeneous choice designs with higher design efficiency. To test this with real data, we conducted a study online on Amazon’s Mechanical Turk or Mturk (see review by Birnbaum 2000 for the use of Mturk in behavioral research) directly contrasting the Sawtooth randomized design with the heterogeneous design from the proposed approach.

5.1. The Empirical Study

Participants in the study were asked to make choices on laptop computers that are meaningful and relevant to the general population online. The laptop computers were characterized by four product attributes each with three possible attribute levels: screen size (12”, 14”, and 15.6”), memory (4 GB, 6 GB, and 8 GB), hard drive (500 GB, 750 GB, and 1 TB), and price (\$749, \$649, and \$549). Data were collected on 240 respondents who were randomly assigned to two groups each of size 120. In each group every respondent was given 15 choice sets with three product alternatives, constructed either by the Sawtooth randomized design (for the 120 respondents in group 1), or the heterogeneous design from the proposed approach (for the 120 respondents in group 2). The same six holdout choice tasks were given to both groups of respondents following the above calibration choice tasks.

The heterogeneous design from the proposed approach and the Sawtooth randomized design used in the empirical study are the same ones that we investigated in the simulation study in §4.2 for the design setting of 15 choice sets. The heterogeneous design from the proposed approach was constructed under the prior specification that $\mu_B = (-1, 0, -1, 0, -1, 0, -1, 0)$ and $\sigma_B = (1, 1, \dots, 1)$. As shown in Table 1, if the true values of μ_B and σ_B were the same as the specified, then the heterogeneous design from the proposed approach would have 99.995% global design efficiency, and the Sawtooth design would have 55.28% global design efficiency.

Note that the attribute levels are based on the effects-type coding as used in previous research (e.g., Arora and Huber 2001; Sándor and Wedel 2002, 2005; Yu et al. 2009), where the part-worths of the three levels of the same attribute sum to 0 and therefore only two parameters are needed for that attribute in the μ_B vector. For example, the first two specified values in the μ_B vector, -1 and 0 , represent the mean part-worths corresponding to the first two levels (12” and 14”) of screen size. This implies that the mean part-worth for the third level (15.6”) of screen size is then 1, the negative sum of the part-worths for the first two levels. Thus, the specified values of μ_B and σ_B make intuitive sense—they indicate that consumers generally prefer larger screens, larger memory, larger hard drive capacity, and lower price, with modest levels of heterogeneity.

Without knowledge of the true values of the model parameters, we cannot use the ERMSEs as defined in §4.4 to compare design efficiency on parameter recovery and prediction. Instead, we report in Table 4 the in-sample model fit and out-of-sample prediction statistics as empirical measures of internal and external validity. The in-sample model fit statistics are based on the calibration choice tasks. We examine both the average log likelihood and the log marginal density (LMD) calculated according to the mean harmonic estimator by Newton and Raftery (1994). The out-of-sample prediction measures are based on the six holdout choice tasks. We examine both the hit rate that measures the predictive validity at the individual level and the mean absolute error (MAE) that measures predictive validity at the aggregate level (Wittink and Bergestuen 1999). Starting with data from group 1 where the choice sets were constructed by the Sawtooth randomized design, we use hierarchical Bayesian analysis and retain 50,000 draws after convergence in the Markov chain Monte Carlo for inferences on the posterior estimates and holdout predictions. We do the same for the data from group 2 where the choice sets were constructed by the heterogeneous design from the proposed approach. Then, based on the 50,000 posterior draws, we calculate the differences in parameter estimates, model fit, and holdout prediction statistics between the two groups.

Table 4 Empirical Test of the Proposed Approach as an Alternative to the Industry Standard

Design	In-sample model fit		Hold-out prediction	
	Avg. log likelihood	LMD	Hit rate (%)	MAE
Proposed approach	-580.91 ^a	-627.93 ^a	83.06 ^a	0.029 ^a
Sawtooth randomized	-691.59	-740.13	79.03	0.039

^aSignificantly different from the corresponding statistics obtained from the Sawtooth randomized approach (α -level = 0.05).

If the 95% highest posterior density (HPD) intervals do not include 0, then the differences are considered to be statistically significant (α -level = 0.05).

The results from the empirical test, as shown in Table 4, confirm the superior performance of the heterogeneous design obtained from the proposed approach in comparison to the Sawtooth randomized design. The significantly higher values in average log likelihood (−580.91 versus −691.59) and LMD (−627.93 versus −740.13) in model fit statistics suggest that data generated by the design from the proposed approach lead to better internal validity in contrast to data generated by the Sawtooth randomized design. In addition, the significantly higher hit rate (83.06% versus 79.03%) and lower MAE (0.029 versus 0.039) suggest that better predictive validity is achieved by using the design from the proposed approach as opposed to the Sawtooth randomized design.

Table 5 shows the posterior parameter estimates of the model parameters μ_B and σ_B from using either the proposed approach or the Sawtooth randomized approach. We find that the two approaches lead to significantly different parameter estimates on two attributes—memory and hard drive capacity. In comparison to the estimates from the Sawtooth approach, the estimates from the proposed approach demonstrate significantly stronger preferences for larger memory and larger hard drive capacity, together with higher levels of heterogeneity. These significant differences in parameter estimates have important implications for substantive marketing problems such as the development of a new product. It makes a strong case for the use of the proposed approach as a valuable alternative to the Sawtooth randomized design. With gains in both model fit and predictive validity, the proposed approach allows managers to make better business decisions based on more accurate estimates of consumer preferences and heterogeneity.

5.2. Difference Between Empirical and Simulated Data

It should be noted that the efficiency gain of the proposed approach over the Sawtooth approach observed in the empirical data is less substantial than that observed in the simulated data. This is similar to findings in Johnson et al. (2005), where the authors compared the adaptive choice-based conjoint (CBC) with the standard CBC designs, both by Sawtooth software. The authors provided an in-depth discussion on possible reasons for this difference between empirical and simulated data. In regards to our study, we believe the following factors are especially relevant: (1) imperfect prior specifications; (2) choice task dependent response errors; and (3) use of simplifying choice decision heuristics. We discuss each factor in some detail next.

Table 5 Posterior Estimates of Model Parameters from the Empirical Data

Attribute	Proposed approach		Sawtooth randomized	
	μ_B	σ_B	μ_B	σ_B
Screen size				
12"	−1.71 (0.21)	1.84 (0.22)	−2.07 (0.20)	1.76 (0.18)
14"	0.17 (0.12)	0.72 (0.12)	0.43 (0.09)	0.58 (0.11)
Memory				
4 GB	−2.39 ^a (0.23)	1.78 ^a (0.20)	−1.47 (0.15)	1.16 (0.13)
6 GB	−0.03 (0.10)	0.41 (0.08)	0.21 (0.08)	0.50 (0.09)
Hard drive				
500 GB	−1.39 ^a (0.17)	1.23 ^a (0.16)	−0.99 (0.12)	0.81 (0.11)
750 GB	0.16 (0.10)	0.49 (0.09)	0.04 (0.08)	0.40 (0.07)
Price				
\$749	−2.39 (0.24)	1.89 (0.21)	−1.97 (0.20)	1.82 (0.19)
\$649	0.35 (0.09)	0.48 (0.09)	0.31 (0.08)	0.48 (0.08)

Note. The posterior standard deviations of the parameter estimates are reported in parentheses.

^aSignificantly different from the corresponding estimate obtained from the Sawtooth randomized approach (α -level = 0.05).

First, in the simulated data we had perfect prior knowledge and incorporated such knowledge in the design construction for the proposed approach. In the empirical data, however, there is no perfect prior knowledge and the prior specifications used for the design construction may deviate from the true values of model parameters (as reflected through the posterior estimates in the empirical data). This deviation decreases the efficiency of the design constructed by the proposed approach and could contribute to a reduction in the efficiency gain of the proposed approach over the Sawtooth approach in the empirical data. Nevertheless, we note that considerable gains in model fit and predictive validity are still attained by the proposed approach. We believe this is partly because the prior specification correctly reflects the rank order of part-worths within each attribute, which worked to the benefit of the proposed approach over the Sawtooth approach.

Second, by accounting for nonzero priors of part-worths and consumer heterogeneity, designs from the proposed approach tend to be more utility balanced (Huber and Zwerina 1996), which increases the difficulty of the choice tasks. In the simulated data, the response errors are generated according to the logit model where the errors are independent of the choice task. However, in reality, responses to more difficult choice tasks may have larger errors, and the increased error may counterbalance the gains in statistical efficiency (Johnson et al. 2003).

Third, in the simulated data, choices are generated according to the compensatory, mixed logit model. However, in reality, respondents may not follow a compensatory evaluation process when making choices and may resort to simplifying choice decision

heuristics (Swait and Adamowicz 2001). For example, respondents may use a two-stage consider-then-choose decision process (e.g., Gilbride and Allenby 2004) as task difficulty increases. It is known that in such cases, designs from the proposed approach, which are optimized for the mixed logit model, may suffer efficiency loss (Liu and Arora 2011).

Regardless of how the above three factors may offset the efficiency gains of the proposed approach, the results from the empirical data show that the proposed approach results in statistically significant gains in model fit and predictive validity. We offer a novel method to construct efficient heterogeneous choice designs. Although it is outside the scope of this paper to decompose the individual effect of these three factors on design efficiency in the field, it is clearly an important avenue for future research.

6. Constraints and Extensions

In this section, we discuss computational constraints of the proposed approach for the situation when there is a large number of attribute and attribute level combinations. We also discuss possible solutions and extensions.

6.1. Computational Constraints and Proposed Solution

The proposed approach stores all possible choice sets in an attempt to find a globally optimal design. As the number of attributes and attribute levels in the choice experiment increases, the number of all possible choice sets increases exponentially. Constrained by the memory limit of the computing platform,⁶ the proposed approach will likely encounter the memory shortage problem in computation. A potential solution to this problem is to reduce the design space for the proposed approach by reducing all possible alternatives to a well-selected subset and then form the choice sets with the subset of alternatives. For example, for a big conjoint study that involves 10 attributes each with three levels, rather than using all possible $3^{10} = 59,049$ product alternatives to form the choice sets, we can reduce the design space significantly by using a subset of 45 alternatives. A larger subset (e.g., 81) can be selected if permitted by the memory capacity of the computing platform.

The subset of alternatives can be selected based on the idea of fractional factorial designs or orthogonal/nearly orthogonal designs (see, for example, Kuhfeld and Tobias 2005). Alternatively we can select

the subset based on the idea of uniform designs (see Fang and Lin 2003, Zhou and Fang 2012). Orthogonal designs strive for one-way and two-way balances over the frequency of occurrences of attribute levels. In other words, they aim for the one- and two-dimensional uniformity over the distribution of attribute levels. In contrast, uniform designs strive for uniformity across all dimensions. As a result, uniform designs have been shown to be more robust against model misspecifications (Fang and Lin 2003). For example, when there are possible interaction effects, uniform designs would allow us to separate the true interaction effects from the main effects, whereas they may be completely confounded in orthogonal designs. Therefore, one could use the uniform designs to select the subset of alternatives for the proposed approach. Using the online catalog of uniform designs publicly available at <http://sites.stat.psu.edu/~rli/DMCE/UniformDesign/>, we obtain the subset of 45 alternatives used to form the reduced design space for the proposed approach. Additional uniform designs can be obtained by using the algorithms discussed in Zhou and Fang (2012).

Table 6 shows the results of comparison between the benchmark designs and the heterogeneous design from the proposed approach with the reduced design space where the choice sets are formed by the subset of 45 alternatives. This reduction effectively resolves the problem of memory shortage for the proposed approach in the case of a big conjoint study that involves 10 attributes each with three levels, four alternatives per choice set, and 15 choice sets per respondent. Whereas there are no significant differences among the designs on the recovery of individual-level part-worths, the efficiency results⁷ based on D-error, predictive accuracy, and recovery of parameters $\theta = (\mu_B, \sigma_B)$ demonstrate the superiority of the heterogeneous design from the proposed approach, even when it is constructed with the reduced design space. In comparison to the separate search approach, the proposed approach with the reduced design space takes only a fraction of the computation time and generates a completely heterogeneous design with small to modest efficiency gains based on three different design efficiency measures. In comparison to the Sawtooth randomized design, the proposed approach with the reduced design space continues to attain substantial and significant gains in efficiency.

These findings demonstrate that one could use the proposed approach with reduced design space

⁶ Currently in Windows-based SAS/IML with which we have coded the proposed approach, the work memory is limited to 2 GB even if there is extra memory available in the computer system. SAS technical support is aware of this issue and is hoping to address it in future releases.

⁷ Because of the memory constraint, we cannot search through all possible choice sets to find a globally optimal continuous design. Therefore, instead of using the global design efficiency measure, we focus on the relative design efficiency measures.

Table 6 Performance of the Proposed Approach with Reduced Design Space vs. Benchmark

	Proposed approach (reduced design space) ($M = 120$)	Separate search approach ($M = 6$)	Sawtooth randomized design ($M = 120$)
Computation time (hh:mm:ss)	00:39:35	02:16:07	00:09:15
Percentage improvement in efficiency	8.52%	—	–65.71%
Efficiency gain in prediction	0.19%	—	–27.40% ^a
Efficiency gain in recovery of $\theta = (\mu_B, \sigma_B)$	1.79% ^a	—	–30.36% ^a
Efficiency gain in recovery of individual-level part-worths	–1.23%	—	–0.91%

Notes. (i) The reduced design space for the proposed approach is formed from a subset of 45 alternatives. (ii) The design setting involves 10 attributes each with three levels; four alternatives per choice set; 15 choice sets per respondent; 120 respondents.

^aSignificantly different from 0 (α -level = 0.05).

for design scenarios that involve a large number of attributes and attribute level combinations. Even for scenarios with a modest number of attributes and attribute levels, the reduction of the design space could be used for the proposed approach if the researcher is willing to sacrifice some design efficiency to achieve faster computation speed. For example, for the design scenarios discussed in §4 that involve four attributes each with three levels, using a reduced design space formed by a subset of 45 alternatives in the proposed approach would reduce the computation time from 2 minutes 15 seconds to 40 seconds, at the cost of 6.52% loss in design efficiency based on the D-error.

In addition to using the uniform designs for robustness against possible model misspecifications in the selection of the subset of alternatives for the proposed approach, further improvement can be made if prior knowledge is available in practical applications. For example, if there is prior knowledge on possible interaction effects, then such knowledge can be incorporated into the definition of the X matrix by including additional columns that represent the possible interaction effects. The selection of the subset of alternatives can then be based on a fractional factorial design that is optimized based on the D-criterion that incorporates the possible interaction effects in the X matrix. Alternatively, a composite design criterion can be used where the composite D-error is the product of the D-error for the main-effect only model and for the possible interaction-effect model. Such composite criterion has been shown to be more robust to misspecifications of interaction effects (Yu et al. 2008). Similarly, when there is prior knowledge that the consumer choice process is likely noncompensatory with screening at the consideration stage, such knowledge can be incorporated by using the design criterion for the two-stage consider-then-choose model (Liu and Arora 2011). Last, if there are prohibitions of certain attribute-level combinations for practical considerations (Chrzan and Orme 2000) then they can be easily accommodated by restricting the subset of alternatives to only those allowed.

6.2. Extension to Other Design Criteria

Beyond the popular D-criterion, the proposed approach easily extends to other design criteria that are based on the Fisher information matrix, such as the A- and V-criteria (Kessels et al. 2006) as well as the M-criterion (Toubia and Hauser 2007). This is because the proposed approach is based on the general equivalence theorem (Kiefer 1974) and the OWEA (Yang et al. 2013), both of which apply to a wide class of design optimality criteria. The expressions of the general equivalence theorem for the A-, V-, and M-criteria are provided in Appendix A. Correspondingly, the formulas for the optimal weight derivations in the proposed approach for the three criteria are provided in Appendix B.

7. Summary and Conclusion

Extant research on choice designs in the marketing literature focuses on the construction of efficient homogeneous designs where every respondent is given the same sets of product alternatives for evaluation. Sándor and Wedel (2005) were the first to propose efficient heterogeneous designs and demonstrate substantial efficiency gain of using heterogeneous designs over homogeneous designs. Although extant computer-search approaches to finding efficient designs work well for the case of homogeneous designs, they become inadequate for the case of heterogeneous designs. The high computation cost significantly hinders the widespread adoption of heterogeneous designs, even when the number of subdesigns for the heterogeneous design is restricted to be small (i.e., six). As a result, randomized designs by Sawtooth software (Chrzan and Orme 2000) are often used in practice where the subdesigns are generated in a balanced but somewhat randomized way fairly quickly but at the expense of design efficiency.

In this paper, we propose a new approach, which not only achieves high efficiency of the resulting heterogeneous choice design but is also computationally feasible. Our approach makes it practical to obtain

Downloaded from informs.org by [128.248.155.225] on 09 April 2016, at 19:45. For personal use only, all rights reserved.

a highly efficient design that is completely heterogeneous with a unique subdesign for each individual respondent in the conjoint choice experiment. The construction of the heterogeneous design occurs all at once before the choice experiment is launched and responses collected. This is different from an adaptive choice design where the choice sets are sequentially constructed and customized on the fly for each respondent based on the respondent's responses to previous questions (e.g., Toubia et al. 2004, 2007).

To the best of our knowledge, our research is the first in the marketing literature to find a completely heterogeneous choice design with assured high global design efficiency. In contrast to extant approaches that rely on the exact design framework, the use of a continuous design framework in the proposed approach allows us to leverage existing mathematical theories to quickly identify a globally optimal continuous design. A rounding procedure that converts the globally optimal continuous design to the corresponding heterogeneous choice design ensures the high global design efficiency of the obtained exact design. In practice, when the number of attributes or attribute levels are moderate, the proposed approach is capable of finding a globally optimal continuous design in the entire design space and allows us to measure the global design efficiency of any given design relative to the globally optimal continuous design. When the number of attributes or attribute levels increase, the design space increases exponentially. In such circumstances, we demonstrate that a reduction of the design space for the proposed approach can be easily accomplished by selecting a subset rather than using all possible product alternatives to form the choice sets in the design space. Our proposed approach is fairly general. When alternative design criteria based on the Fisher information matrix are of interest, including the A-, V-, and M-criteria, the proposed approach can be easily extended as well.

In conclusion, our proposed approach provides a practical and valuable tool to marketing practitioners. We believe that using this approach to generate efficient heterogeneous designs can benefit marketing researchers/managers in many areas, including but not limited to new product development, sales force management, pricing, etc. First, the significant gain in design efficiency in comparison to designs obtained from extant approaches allows managers to estimate consumer preferences more accurately and subsequently make better decisions based on these estimates. Second, the fast computation speed of the proposed approach makes it practical for generating a completely heterogeneous design at the individual level so that there is a unique subdesign for each individual respondent. We hope this research will generate more interest in the adoption

of heterogeneous designs and stimulate more interest in further research in the area of efficient choice designs.

Supplemental Material

Supplemental material to this paper is available at <http://dx.doi.org/10.1287/mksc.2014.0897>.

Acknowledgments

The authors thank the editor, associate editor, and two anonymous reviewers for their valuable suggestions that greatly improved the paper. The authors also thank Neeraj Arora and Jan Heide for helpful comments, Bryan Orme for access to the Sawtooth software, and Vincent Chan for programming support. This research was supported in part by the Graduate School Research award and the Vilas Life Cycle Professorship award from the University of Wisconsin–Madison, the Managerial Studies Research and Development Fund from the University of Illinois at Chicago, and the Donald W. Reynolds Journalism Institute. The order of authorship is alphabetical to reflect equal contribution.

Appendix A

General Equivalence Theorem Expressions for the A-, V-, and M-Criteria

Note that the A-criterion is to minimize the trace of the inverse of the Fisher information matrix, that is, $\text{Tr}(I^{-1}(X))$. The M-criterion is to minimize $\text{Tr}(MI^{-1}(X)M')$, which is equivalent to minimizing $\text{Tr}(M'MI^{-1}(X))$. The V-criterion is to minimize average prediction variances over the choice probabilities of product alternatives in all Q possible future choice sets, that is, to minimize $\sum_{q=1}^Q \sum_{j=1}^J Z'(x_{qj}) \cdot I^{-1}(X)Z(x_{qj}) = \text{Tr}\{[\sum_{q=1}^Q \sum_{j=1}^J Z(x_{qj})Z'(x_{qj})]I^{-1}(X)\}$, where $Z(x_{qj}) = \partial(\hat{p}_{qj}(x_{qj}, \theta))/\partial\theta$, and $\hat{p}_{qj}(x_{qj}, \theta)$ denotes the predicted choice probability for alternative j in choice set q . Thus, the A-, V-, and M-criteria can be expressed in a general form as $\text{Tr}(PI^{-1}(X))$, where $P = I$ for the A-criterion, $P = \sum_{q=1}^Q \sum_{j=1}^J Z(x_{qj})Z'(x_{qj})$ for the V-criterion, and $P = M'M$ for the M-criterion. The general equivalence theorem for the three criteria can then be expressed as follows.

General Equivalence Theorem. A design \tilde{X}^* is globally optimal in the continuous design space if and only if

$$d(C_k, \tilde{X}^*) = \text{Tr}\{PI(\tilde{X}^*)^{-1}[I(C_k) - I(\tilde{X}^*)]I(\tilde{X}^*)^{-1}\} \leq 0,$$

for any C_k that belongs to the space of all possible choice sets.

Here, $I(C_k)$ is the Fisher information matrix for choice set C_k , $I(\tilde{X}^*)$ is the Fisher information matrix for the continuous design \tilde{X}^* , $P = I$ for the A-criterion, $P = \sum_{q=1}^Q \sum_{j=1}^J Z(x_{qj})Z'(x_{qj})$ for the V-criterion, and $P = M'M$ for the M-criterion.

Appendix B

Procedure to Derive Optimal Weights Using Newton's Iteration Method

Suppose there are r choice sets in the initial design, and let $\{w_1, \dots, w_{r-1}, w_r\}$ be the associated initial weights. Define $W^{(t)} = (w_1, \dots, w_{r-1})$ and $\alpha = 1$. Notice that $w_r = 1 - \sum_{i=1}^{r-1} w_i$. So we only need to update $w^{(t)}$, the first $r - 1$ weights after the t th iteration.

a. First, $W^{(t)}$ can be updated using the following formula

$$W^{(t)} = \begin{cases} W^{(t-1)} - \alpha \left(\frac{\partial^2 \log |\tilde{I}|}{\partial W \partial W'} \Big|_{W=W^{(t-1)}} \right)^{-1} \frac{\partial \log |\tilde{I}|}{\partial W} \Big|_{W=W^{(t-1)}}, \\ \text{D-optimal} \\ W^{(t-1)} - \alpha \left(\frac{\partial^2 \text{Tr}(\tilde{P} \tilde{I}^{-1})}{\partial W \partial W'} \Big|_{W=W^{(t-1)}} \right)^{-1} \frac{\partial \text{Tr}(\tilde{P} \tilde{I}^{-1})}{\partial W} \Big|_{W=W^{(t-1)}}, \\ \text{A, V, M-optimal.} \end{cases}$$

Here, \tilde{I} is the Fisher information matrix for the initial design with weights W . $P = I$ for the A-criterion, $P = \sum_{q=1}^Q \sum_{j=1}^J Z(x_{qj})Z'(x_{qj})$ for the V-criterion, and $P = M'M$ for the M-criterion.

b. Check if there are nonpositive components of $W^{(t)}$. If so, go to step (c2), otherwise proceed to c1.

c1. Check if $\|\partial \log |\tilde{I}| / \partial W|_{W=W^{(t-1)}}\|$ is less than a prespecified small positive value, say, 10^{-6} , for the D-criterion. Similarly, check $\|(\partial \text{Tr}(\tilde{P} \tilde{I}^{-1})) / \partial W|_{W=W^{(t)}}\|$ for A-, V-, and M-criterion. If so, $W^{(t)}$ is the vector of optimal weights. Otherwise, start the next iteration.

c2. Reduce α to $\alpha/2$. Repeat (a) and (b) until α reaches a prespecified value, say 10^{-6} .

Remove the choice set with the smallest weight, and then go to (a) with the new set of choice sets and update their weights.

Appendix C

Proof of the Assured High Efficiency of the Resulting Exact Design

Assume the globally optimal continuous design is $\tilde{X}^* = \{(C_k, w_k)\}, k = 1, \dots, K$ with the maximum weight less than or equal to $1/S$. The detailed steps of the procedure that converts the globally optimal continuous design to an exact design are as follows:

1. Rank order the choice sets within the globally optimal continuous design by their corresponding weights w_k (from high to low) multiplied by the number of respondents and the number of choice sets per respondent ($\tilde{l}_{0k} = HS w_k$). Assign the first S choice sets to the first respondent ($h = 1$).

2. For the S choice sets that have been assigned in the previous step, let $\tilde{l}_{hk} = \tilde{l}_{h-1,k} - 1$. For the remaining choice sets, let $\tilde{l}_{hk} = \tilde{l}_{h-1,k}$. Reorder the choice sets according to \tilde{l}_{hk} and assign the first S choice sets to the next respondent (i.e., respondent $h + 1$).

3. Repeat step 2 until we get S choice sets for each of the H respondents.

Let \tilde{n}_k be the total number of replications of choice sets C_k in the resulting exact design. For the proof of the assured high efficiency of the resulting exact design, we need to show that

$$[HS w_k] \leq \tilde{n}_k \leq \lceil HS w_k \rceil, \quad k = 1, \dots, K. \quad (C1)$$

Here, $[HS w_k]$ means the largest integer less than or equal to $HS w_k$ and $\lceil HS w_k \rceil$ means the smallest integer greater than or equal to $HS w_k$. The above inequality ensures that \tilde{n}_k/HS , the weight of each choice set of the continuous design corresponding to the resulting exact design, is very close to the weight w_k in the globally optimal continuous design. As a

result, the resulting exact design is assured to have high global efficiency.

First, we show that the left half of (C1) (i.e., $[HS w_k] \leq \tilde{n}_k$) holds through proof by contradiction. Suppose that $\tilde{n}_k < [HS w_k]$ for some k , say, k_1 , which is equivalent to $\tilde{n}_{k_1} \leq [HS w_{k_1}] - 1$ since they are integers. According to the procedure, \tilde{l}_{hk_1} starts with $\tilde{l}_{0k_1} = HS w_{k_1}$ and it remains the same if the choice set C_{k_1} is not chosen, or \tilde{l}_{hk_1} is reduced by 1 each time the choice set C_{k_1} is chosen, as h moves from 1 to H . As a result

$$\tilde{l}_{Hk_1} \geq \tilde{l}_{0k_1} - \tilde{n}_{k_1} \geq HS w_{k_1} - ([HS w_{k_1}] - 1) \geq 1. \quad (C2)$$

On the other hand, all $\tilde{l}_{Hk} \geq 0$ for all k . This must hold true according to the following proof by contradiction. Let $k_2 = \arg \min_k \tilde{l}_{Hk}$. Suppose that $\tilde{l}_{Hk_2} < 0$. Then choice set C_{k_2} must be chosen for the last respondent (i.e., respondent H), and thus $\tilde{l}_{H-1,k_2} = \tilde{l}_{Hk_2} + 1$. Otherwise, the S th largest $\tilde{l}_{H-1,k}$ must be greater than $\tilde{l}_{H-1,k_2} + 1$ by the definition of k_2 and the nature of the procedure. This would consequently mean that \tilde{l}_{H-2,k_2} is not among the S largest $\tilde{l}_{H-2,k}, k = 1, \dots, K$ because $\tilde{l}_{H-2,k} \geq \tilde{l}_{H-1,k}$ and therefore there are at least S $\tilde{l}_{H-2,k}$'s larger than $\tilde{l}_{H-1,k_2} + 1$, which is larger than or equal to \tilde{l}_{H-2,k_2} by the definition of the procedure. This would mean that choice set C_{k_2} is not chosen for respondent $H - 1$ and thus $\tilde{l}_{H-2,k_2} = \tilde{l}_{H-1,k_2}$. We can continue the same argument until \tilde{l}_{0,k_2} , and we have $\tilde{l}_{0,k_2} = \tilde{l}_{Hk_2} < 0$, which contradicts to the condition that $w_k > 0$ for all $k = 1, \dots, K$. This contradiction proves that choice set C_{k_2} must be chosen for the last respondent (i.e., respondent H), which means that \tilde{l}_{H-1,k_2} is among the S largest $\tilde{l}_{H-1,k}, k = 1, \dots, K$. Next we prove that \tilde{l}_{Hk_2} cannot be less than 0. If $\tilde{l}_{Hk_2} < 0$ then we have $\tilde{l}_{H-1,k_2} = \tilde{l}_{Hk_2} + 1 < 1$. This would suggest that choice set C_{k_1} must have also been chosen for the last respondent because $\tilde{l}_{H-1,k_1} \geq \tilde{l}_{Hk_1} \geq 1 > \tilde{l}_{Hk_2} + 1 = \tilde{l}_{H-1,k_2}$, which means that \tilde{l}_{H-1,k_1} must be among the S largest $\tilde{l}_{H-1,k}, k = 1, \dots, K$. Therefore, we have $\tilde{l}_{H-1,k_1} = \tilde{l}_{Hk_1} + 1 > \tilde{l}_{H-1,k_2} + 1$, which means that \tilde{l}_{H-1,k_1} is greater than the S th largest $\tilde{l}_{H-1,k}$ by more than 1. Subsequently, choice set C_{k_1} must have been chosen for the respondent $H - 1$ too because $\tilde{l}_{H-1,k} \leq \tilde{l}_{H-2,k} \leq \tilde{l}_{H-1,k} + 1$ and thus \tilde{l}_{H-2,k_1} must still be among the S largest $\tilde{l}_{H-2,k}$. Therefore we have $\tilde{l}_{H-2,k_1} = \tilde{l}_{H-1,k_1} + 1 = \tilde{l}_{Hk_1} + 2$. We can continue this argument until \tilde{l}_{0,k_1} , and we have $\tilde{l}_{0,k_1} = \tilde{l}_{Hk_1} + H \geq H + 1$, which contradicts to the assumption that $\max_{1 \leq k \leq K} w_k \leq 1/S$. This contradiction proves that it must be that all $\tilde{l}_{Hk} \geq 0$ for all k .

Now that we have all $\tilde{l}_{Hk} \geq 0$ for all k and $\tilde{l}_{Hk_1} \geq 1$. By the definition of exact design for H respondents each with S choice sets, we have $\sum_{k=1}^K \tilde{n}_k = HS$. Thus we have

$$\sum_{k=1}^K (\tilde{l}_{Hk} + \tilde{n}_k) \geq 1 + HS, \quad (C3)$$

which contradicts to the fact that

$$\sum_{k=1}^K (\tilde{l}_{Hk} + \tilde{n}_k) = \sum_{k=1}^K \tilde{l}_{0k} = HS. \quad (C4)$$

This contradiction proves that the left half of (C1) holds, that is, $\tilde{n}_k \geq [HS w_k]$.

Now let us prove the right half of Inequality (C1). Suppose that there exists k_3 , such that $\tilde{n}_{k_3} > \lceil HS w_{k_3} \rceil$, which

is equivalent to $\bar{n}_{k_3} \geq \lceil HS w_{k_3} \rceil + 1$. Similar to (C2), we must have $\tilde{l}_{Hk_3} < -1$. Let $k_5 = \arg \min_k \tilde{l}_{Hk}$ and it must be that $\tilde{l}_{Hk_5} < -1$. Let $k_4 = \arg \max_k \tilde{l}_{Hk}$. We must have $\tilde{l}_{Hk_4} > 0$. Otherwise, by the similar argument in (C3), we would have

$$\sum_{k=1}^K (\tilde{l}_{Hk} + \bar{n}_k) \leq HS - 1, \quad (C5)$$

which contradicts to (C4).

Next, with $\tilde{l}_{Hk_4} > 0$ and $\tilde{l}_{Hk_5} < -1$, we have two possible outcomes: (i) we can have $\tilde{l}_{0, k_5} = \tilde{l}_{Hk_5}$ using exactly the same argument as that for $\tilde{l}_{0, k_2} = \tilde{l}_{Hk_2}$ or (ii) $\tilde{l}_{0, k_4} = \tilde{l}_{Hk_4} + H > H$ using exactly the same argument as that for $\tilde{l}_{0, k_1} = \tilde{l}_{Hk_1} + H \geq H + 1$. The first outcome contradicts with the condition $w_k > 0$ for all $k = 1, \dots, K$, and the second outcome contradicts with the assumption that $\max_{1 \leq k \leq K} w_k \leq 1/S$. This completes the proof.

References

- Arora N, Huber J (2001) Improving parameter estimates and model prediction by homogeneous customization in choice experiments. *J. Consumer Res.* 28(2):273–283.
- Atkinson A, Donev A, Tobias R (2007) *Optimum Experimental Designs, with SAS* (Oxford University Press, New York).
- Ben-Akiva M, Lerman S (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand* (MIT Press, Cambridge, MA).
- Birnbaum MH (2000) *Psychological Experiments on the Internet* (Academic Press, San Diego).
- Chaloner K, Verdinelli I (1995) Bayesian experimental design: A review. *Statist. Sci.* 10(3):273–304.
- Chrzan K, Orme B (2000) An overview and comparison of design strategies for choice-based conjoint analysis. Technical report, Sawtooth software. <http://www.sawtoothsoftware.com/download/techpap/desgnbc.pdf>.
- Ding M (2007) An incentive-aligned mechanism for conjoint analysis. *J. Marketing Res.* 44(2):214–223.
- Ding M, Grewal R, Liechty J (2005) Incentive-aligned conjoint analysis. *J. Marketing Res.* 42(1):67–82.
- Ding M, Park YH, Bradlow ET (2009) Barter markets for conjoint analysis. *Management Sci.* 55(6):1003–1017.
- Fang KT, Lin DKJ (2003) Uniform experimental design and its applications in industry. Rao CR, Khattree R, eds. *Handbook of Statistics in Industry*, Vol. 22 (North Holland, Amsterdam), 131–170.
- Fedorov VV, Hackl P (1997) *Model-Oriented Design of Experiments* (Springer Verlag, New York).
- Gilbride TJ, Allenby GM (2004) A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Sci.* 23(3):391–406.
- Gilbride TJ, Allenby GM (2006) Estimating heterogeneous EBA and economic screening rule choice models. *Marketing Sci.* 25(5):494–509.
- Green PE, Srinivasan V (1990) Conjoint analysis in marketing: New developments with implications for research and practice. *J. Marketing* 54(4):3–19.
- Haaïjer R, Kamakura WA, Wedel M (2001) The no-choice alternative in conjoint choice experiments. *Internat. J. Market Res.* 43(1):93–106.
- Huber J, Zwerina K (1996) The importance of utility balance in efficient choice designs. *J. Marketing Res.* 33(3):307–317.
- Johnson R, Huber J, Bacon L (2003) Adaptive choice-based conjoint designs. *Sawtooth Software Conf. Proc.* 333–343. <http://www.sawtoothsoftware.com/support/technical-papers/conference-proceedings>.
- Johnson R, Orme B, Huber J, Pinnell J (2005) Testing adaptive choice-based conjoint designs. Technical paper, Sawtooth software. <http://www.sawtoothsoftware.com/download/techpap/acbc3.pdf>.
- Kessels R, Goos P, Vandebroek M (2006) A comparison of criteria to design efficient choice experiments. *J. Marketing Res.* 43(3):409–419.
- Kessels R, Jones B, Goos P, Vanderbroek M (2009) An efficient algorithm for constructing Bayesian optimal choice designs. *J. Bus. Econom. Statist.* 27(2):279–291.
- Kiefer J (1974) General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* 2(5):849–879.
- Kuhfeld WF, Tobias R (2005) Large factorial designs for product engineering and marketing research applications. *Technometrics* 47(2):132–141.
- Kuhfeld WF, Tobias RD, Garratt M (1994) Efficient experimental design with marketing research applications. *J. Marketing Res.* 31(4):545–557.
- Liu Q, Arora N (2011) Efficient choice designs for a consider-then-choose model. *Marketing Sci.* 30(2):321–338.
- Liu Q, Dean AM, Allenby GM (2012) Bayesian designs for hierarchical linear models. *Statistica Sinica* 22(1):393–417.
- McFadden D (1974) Conditional logit analysis of qualitative choice behavior. Zarembka P, ed. *Frontiers in Econometrics* (Academic Press, New York), 105–142.
- Meyer R, Nachtsheim C (1995) The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* 37(1):60–69.
- Netzer O, Toubia O, Bradlow ET, Dahan E, Evgeniou T, Feinberg FM, Feit EM, et al. (2008) Beyond conjoint analysis: Advances in preference measurement. *Marketing Lett.* 19(3–4):337–354.
- Newton M, Raftery A (1994) Approximate Bayesian inference with weighted likelihood bootstrap. *J. Royal Statist. Soc. Series B* 56(1):3–48.
- Pukelsheim F (1993) *Optimal Design of Experiments* (John Wiley, New York).
- Sándor Z, Wedel M (2001) Designing conjoint choice experiments using managers' prior beliefs. *J. Marketing Res.* 38(4):430–444.
- Sándor Z, Wedel M (2002) Alternative construction in experimental choice designs for mixed logit models. *Marketing Sci.* 21(4):455–475.
- Sándor Z, Wedel M (2005) Heterogeneous conjoint choice designs. *J. Marketing Res.* 42(2):210–218.
- Silvey SD (1980) *Optimal Design* (Chapman and Hall, London).
- Swait J, Adamowicz W (2001) The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *J. Consumer Res.* 28(1):135–148.
- Toubia O, Hauser JR (2007) On managerial efficiency for experimental designs. *Marketing Sci.* 26(6):851–858.
- Toubia O, Hauser JR, Garcia R (2007) Probabilistic polyhedral methods for adaptive choice-based conjoint analysis: Theory and application. *Marketing Sci.* 26(5):596–610.
- Toubia O, Hauser JR, Simester DI (2004) Polyhedral methods for adaptive choice-based conjoint analysis. *J. Marketing Res.* 41(1):116–131.
- Toubia O, de Jong MG, Stieger D, Fueller J (2012) Measuring consumer preferences using conjoint poker. *Marketing Sci.* 31(1):138–156.
- Wittink DR, Bergstuen T (1999) Forecasting with conjoint analysis. Armstrong JS, ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Kluwer Academic, Norwell, MA), 147–167.
- Yang M, Stufken J (2009) Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.* 37(1):518–541.
- Yang M, Biedermann S, Tang E (2013) On optimal designs for nonlinear models: A general and efficient algorithm. *J. Amer. Statist. Assoc.* 108(504):1411–1420.
- Yu J, Goos P, Vandebroek M (2008) Model-robust design of conjoint choice experiments. *Comm. Statist.: Simulation Comput.* 37(8):1603–1621.
- Yu J, Goos P, Vanderbroek M (2009) Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Sci.* 28(1):122–135.
- Zhou YD, Fang KT (2012) An efficient method for constructing uniform designs with large size. *Comput. Statist.* 28(3):1319–1331.