MARKETING SCIENCE
Vol. 28, No. 1, January–February 2009, pp. 122–135
ISSN 0732-2399 | EISSN 1526-548X | 09 | 2801 | 0122

DOI 10.1287/mksc.1080.0386


# Efficient Conjoint Choice Designs in the Presence of Respondent Heterogeneity

Jie Yu
Faculty of Business and Economics, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium,
jie.yu@econ.kuleuven.be

Peter Goos
Faculty of Applied Economics, Universiteit Antwerpen, B-2000 Antwerpen, Belgium, peter.goos@ua.ac.be

Martina Vandebroek
Faculty of Business and Economics and Leuven Statistics Research Centre, Katholieke Universiteit Leuven,
B-3000 Leuven, Belgium, martina.vandebroek@econ.kuleuven.be

Random effects or mixed logit models are often used to model differences in consumer preferences. Data from choice experiments are needed to estimate the mean vector and the variances of the multivariate heterogeneity distribution involved. In this paper, an efficient algorithm is proposed to construct semi-Bayesian *D*-optimal mixed logit designs that take into account the uncertainty about the mean vector of the distribution. These designs are compared to locally *D*-optimal mixed logit designs, Bayesian and locally *D*-optimal designs for the multinomial logit model and to nearly orthogonal designs (Sawtooth (CBC)) for a wide range of parameter values. It is found that the semi-Bayesian mixed logit designs outperform the competing designs not only in terms of estimation efficiency but also in terms of prediction accuracy. In particular, it is shown that assuming large prior values for the variance parameters for constructing semi-Bayesian mixed logit designs is most robust against the misspecification of the prior mean vector. In addition, the semi-Bayesian mixed logit designs are compared to the fully Bayesian mixed logit designs, which take also into account the uncertainty about the variances in the heterogeneity distribution and which can be constructed only using prohibitively large computing power. The differences in estimation and prediction accuracy turn out to be rather small in most cases, which indicates that the semi-Bayesian approach is currently the most appropriate one if one needs to estimate mixed logit models.

*Key words*: semi-Bayesian mixed logit design; fully Bayesian mixed logit design; heterogeneity; prediction accuracy; multinomial logit design; model-robust design; *D*-optimality; design construction algorithm
*History*: Received: March 29, 2007; accepted: January 9, 2008; processed by Michel Wedel. Published online in *Articles in Advance* November 5, 2008.

## 1. Introduction

In marketing, conjoint choice experiments have become popular to explore consumer preferences for certain characteristics of products or services. The data from such experiments are often analyzed by a multinomial logit model (McFadden 1974). The major advantage of this model is its simple form for choice probabilities. However, this simple model has several shortcomings. One of the main shortcomings is that it does not take into account the heterogeneity in consumer or respondent preferences. Incorporating consumer heterogeneity when analyzing consumer behavior is an important topic in the recent marketing literature (Allenby et al. 1998, Arora and Henderson 2007, Luo et al. 2007, Sándor and Wedel 2002, Wedel et al. 1999).

Several models that can deal with respondent heterogeneity have been explored. A latent class model assumes that the individual parameters are drawn from a discrete distribution and is appropriate when several homogeneous groups of respondents can be extracted from a heterogeneous group of data. However, when the true representation of heterogeneity is continuous, the latent class model might not be able to provide accurate estimates (Allenby et al. 1998, Sándor and Wedel 2002). In this situation, models like the probit (Haaijer et al. 1998) or the mixed logit model (McFadden and Train 2000; Revelt and Train 1998; Sándor and Wedel 2002, 2005), which assume that the coefficients are drawn from a continuous distribution, are preferred over the latent class model. The probit model requires normal distributions for all unobserved components of utility (Train 2003). However, in some cases, other distributions are more appropriate than the normal distribution. In addition to that, the popularity of the probit model is limited because of its complicated structure and high computation time (Sándor and Wedel 2002). The mixed logit model is a highly flexible discrete-choice model that can model the heterogeneity of the respondents in a very general way because it is not restricted to normal distributions. It is essentially a multinomial logit model with coefficients that follow a distribution across respondents. McFadden and Train (2000) show that any discrete-choice model can be approximated to any degree of accuracy by a mixed logit model. Compared to the probit model, the simulation of choice probabilities is computationally simpler for the mixed logit model. Because of all of these reasons, the estimation of the mixed logit model has received considerable interest in the literature recently (Huber and Train 2001, Revelt and Train 1998, Train 2003).

A great challenge for the mixed logit model is the quality of the data. Hensher and Greene (2003) state that estimating the mixed logit model certainly demands higher-quality data than the multinomial logit model because it offers an extended framework within which a greater amount of true behavioral variability in choice making can be captured. This implies that the data collection becomes especially important when the goal is to fit a mixed logit model. To avoid situations where the data do not contain enough information for an efficient parameter estimation, it is crucial to search for efficient experimental designs.

In the literature on the optimal design of choice experiments, the focus has mainly been on designs for the multinomial logit model, thereby ignoring the heterogenous preferences across respondents (Huber and Zwerina 1996, Kessels et al. 2006, Sándor and Wedel 2001). The importance of incorporating respondent heterogeneity in design construction was demonstrated by Sándor and Wedel (2002). They showed the advantage of using a mixed logit design over the multinomial logit design. In their locally optimal design procedure, they assume values for the mean partworths, denoted by $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ in this article, and for the covariances $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ of the individual-level coefficients. As a result, they ignore the uncertainty about these values and treat the unknown model parameters as known when constructing the design. To take into account the uncertainty about the unknown model parameters, the most natural approach is the Bayesian approach. It was introduced in the context of conjoint choice design by Sándor and Wedel (2001) for constructing designs for multinomial logit models. They showed that Bayesian designs perform better than classical locally *D*-optimal designs under almost all of the conditions that they have studied. Only when the researcher has reliable knowledge about the true values of the unknown parameters does the locally *D*-optimal design perform better. In this paper, we extend their work to mixed logit models.

However, as shown by Sándor and Wedel (2002), determining the information matrix for the mixed logit model involves a computationally intensive numerical integration over the distribution of the random model coefficients. This makes the construction of Bayesian designs for that model with large numbers of attributes and attribute levels a real challenge. In this paper, we introduce a fast and efficient algorithm that reduces the computation time dramatically. We consider a semi-Bayesian approach, which takes into account the uncertainty about the mean parameter $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ but not about the covariances $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, and a fully Bayesian approach, which takes into account the uncertainty about both $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$. The fully Bayesian mixed logit design is very computationally demanding even with our proposed algorithm. For a moderately sized design problem such as a design with 12 choice sets of three alternatives described by four attributes each at three levels, a single computer needs more than six months to finish 1,000 runs of the algorithm. However, more than this number of runs is needed to obtain an optimal fully Bayesian mixed logit design. The large computation time for generating fully Bayesian mixed logit designs makes the fully Bayesian approach infeasible for practical application. Therefore, we dedicate ourselves to studying the performance of the semi-Bayesian mixed logit designs, which are fast to compute by our algorithm. However, we used a large number of supercomputers to do the computations to be able to compare the semi-Bayesian with the fully Bayesian mixed logit designs. The fully Bayesian designs we report on are generated using 1,500 runs of our algorithm. In §6, we show that the differences in estimation efficiency and in prediction accuracy between the fully Bayesian and the semi-Bayesian mixed logit designs are generally rather small. It is also shown that the semi-Bayesian mixed logit designs perform well for a large range of mean and variance parameters.

The main focus of this paper is to evaluate the advantages of the semi-Bayesian *D*-optimal designs for the mixed logit model over the locally *D*-optimal designs for that model in terms of estimation efficiency and predictive accuracy under various conditions and to study the sensitivity of the semi-Bayesian mixed logit designs to the misspecification of $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$. In addition, we examine how well the Bayesian and the locally *D*-optimal designs for the multinomial logit model, which ignores the respondent heterogeneity, and a nearly orthogonal design generated using Sawtooth (CBC) perform when respondents are heterogeneous.

In the next section, we discuss the structure of a mixed logit model and introduce the design selection criterion utilized in this article. In §3, we present the algorithm to construct efficient designs for the mixed logit model in a computationally efficient way. In §4, we describe the details of the simulation study and the performance criteria. In §5, we evaluate the semi-Bayesian designs in terms of the efficiency of parameter estimation and of the predictive performance and, in §6, we compare the fully Bayesian to the semi-Bayesian designs with respect to these criteria. Section 7 contains a summary of the main findings.

## 2. Mixed Logit Designs

### 2.1. Mixed Logit Model

Mixed logit probabilities are integrals of the standard multinomial logit probabilities over a density function $f(\boldsymbol{\beta})$ for the parameters (Train 2003). The probability that profile $k$ is chosen from choice set $s$ is therefore

$$\pi_{ks} = \int p_{ks}(\boldsymbol{\beta}) f(\boldsymbol{\beta})\, d\boldsymbol{\beta}, \tag{1}$$

where $p_{ks}(\boldsymbol{\beta})$ is the multinomial logit probability evaluated at the parameter values contained in $\boldsymbol{\beta}$:

$$p_{ks}(\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}'_{ks}\boldsymbol{\beta})}{\sum_{i=1}^{K} \exp(\mathbf{x}'_{is}\boldsymbol{\beta})}, \tag{2}$$

with $K$ the number of profiles in each choice set, $\mathbf{x}_{ks}$ a $p$-dimensional vector characterizing the attributes of profile $k$ in choice set $s$, and $\boldsymbol{\beta}$ a $p$-dimensional coefficient vector containing the effects of the different attribute levels on the utility.

To capture the heterogeneity among respondent preferences, we use the same assumption as Sándor and Wedel (2002) that the heterogeneity distribution $f(\boldsymbol{\beta})$ is a multivariate normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and diagonal covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$. We call $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = (\sigma_1, \sigma_2, \ldots, \sigma_p)'$ the heterogeneity vector because it captures the heterogeneity across respondents. The larger the values in $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$, the larger the degree of heterogeneity among respondents. Now, $\boldsymbol{\beta}$ can be written as $\boldsymbol{\beta} = \boldsymbol{\mu}_{\boldsymbol{\beta}} + \mathbf{V}\boldsymbol{\sigma}_{\boldsymbol{\beta}}$, where $\mathbf{V}$ is a diagonal matrix having the random vector $\mathbf{v} = (v_1, v_2, \ldots, v_p)'$ with independent standard normal elements on its diagonal. Substituting this expression in (2), the logit probabilities become functions of the random vector $\mathbf{v}$ and the mixed logit probabilities in (1) can be expressed as

$$\pi_{ks} = \int p_{ks}(\mathbf{v}) h(v_1) h(v_2) \cdots h(v_p)\, d\mathbf{v}, \tag{3}$$

where $h$ represents the standard normal density function. In this article, we are interested in estimating the mean parameter vector $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and the heterogeneity vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ efficiently. In the next section, we introduce a design criterion that can be used to select a design that guarantees an efficient estimation of the $2p$ parameters contained in $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$.

### 2.2. Design Efficiency Criterion

A well-known criterion for evaluating the efficiency of experimental designs is the $D$-optimality criterion.[1] It is based on the determinant of the information matrix on the unknown model parameters, which, for the mixed logit model, are contained within the vectors $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$. The information matrix on $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$, which is inversely proportional to the covariance matrix of the parameter estimates, is given by the $2p \times 2p$ dimensional matrix

$$\mathcal{I}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}} \mid \mathbf{X}) = N \sum_{s=1}^{S} \begin{bmatrix} \mathbf{A}'_s \boldsymbol{\Pi}_s^{-1} \mathbf{A}_s & \mathbf{A}'_s \boldsymbol{\Pi}_s^{-1} \mathbf{B}_s \\ \mathbf{B}'_s \boldsymbol{\Pi}_s^{-1} \mathbf{A}_s & \mathbf{B}'_s \boldsymbol{\Pi}_s^{-1} \mathbf{B}_s \end{bmatrix}, \tag{4}$$

where

$$\mathbf{A}_s = \int [\mathbf{P}_s(\mathbf{v}) - \mathbf{p}_s(\mathbf{v})\mathbf{p}'_s(\mathbf{v})] \mathbf{X}_s h(v_1) h(v_2) \cdots h(v_p)\, d\mathbf{v}, \tag{5}$$

$$\mathbf{B}_s = \int [\mathbf{P}_s(\mathbf{v}) - \mathbf{p}_s(\mathbf{v})\mathbf{p}'_s(\mathbf{v})] \mathbf{X}_s \mathbf{V} h(v_1) h(v_2) \cdots h(v_p)\, d\mathbf{v}, \tag{6}$$

$N$ is the number of respondents, $S$ is the number of choice sets, $\mathbf{p}_s(\mathbf{v}) = [p_{1s}(\mathbf{v}), p_{2s}(\mathbf{v}), \ldots, p_{Ks}(\mathbf{v})]'$, $\mathbf{P}_s(\mathbf{v}) = \mathrm{diag}(p_{1s}(\mathbf{v}), p_{2s}(\mathbf{v}), \ldots, p_{Ks}(\mathbf{v}))$, $\mathbf{X}$ is the entire design matrix, $\mathbf{X}_s$ is the design matrix for choice set $s$, and $\boldsymbol{\Pi}_s = \mathrm{diag}(\pi_{1s}, \ldots, \pi_{Ks})$. This expression was derived by Sándor and Wedel (2002), who constructed locally $D$-optimal designs for the mixed logit model by minimizing

$$D_M\text{-error} = \det\{\mathcal{I}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}} \mid \mathbf{X})^{-1}\}^{1/2p}. \tag{7}$$

The semi-Bayesian mixed (SBM) logit designs introduced in this paper were constructed by minimizing

$$D_{\mathrm{SBM}}\text{-error} = \int \det\{\mathcal{I}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}} \mid \mathbf{X})^{-1}\}^{1/2p} g(\boldsymbol{\mu}_{\boldsymbol{\beta}})\, d\boldsymbol{\mu}_{\boldsymbol{\beta}}, \tag{8}$$

where $g(\boldsymbol{\mu}_{\boldsymbol{\beta}})$ is the prior distribution for the mean parameter vector $\boldsymbol{\mu}_{\boldsymbol{\beta}}$. This distribution can be informative or uninformative depending on the amount of information available to the researcher.

The fully Bayesian mixed (BM) logit design used in §6 was constructed by minimizing

$$D_{\mathrm{BM}}\text{-error} = \int \det\{\mathcal{I}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}} \mid \mathbf{X})^{-1}\}^{1/2p} \cdot q(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}})\, d\boldsymbol{\mu}_{\boldsymbol{\beta}}\, d\boldsymbol{\sigma}_{\boldsymbol{\beta}}, \tag{9}$$

where $q(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}})$ is the joint distribution of $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$. If $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ are assumed independent, then $q(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}}) = g(\boldsymbol{\mu}_{\boldsymbol{\beta}}) m(\boldsymbol{\sigma}_{\boldsymbol{\beta}})$, where $m(\boldsymbol{\sigma}_{\boldsymbol{\beta}})$ is the prior distribution of the heterogeneity parameter $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$.

[1] A comparison of $D$, $A$, $G$, and $V$ criteria for conjoint choice designs was given by Kessels et al. (2006). They show that $D$-optimal designs are nearly as good as the $G$- and $V$-optimal designs in terms of prediction quality but much faster to compute. Recently, Toubia and Hauser (2007) adapted the $D$ and $A$ criteria to focus more on the quantities of managerial interest.

# 3. Design Construction Algorithm

As mentioned by Sándor and Wedel (2002), the large computation time makes it infeasible to construct Bayesian designs for realistic problems using their RSC algorithm (relabeling, sampling, and cycling). In this section, we introduce a design algorithm, which we call the semi-Bayesian algorithm (SB), for generating semi-Bayesian $D$-optimal mixed logit designs. This algorithm allows the construction of large designs for mixed logit models and the incorporation of the uncertainty about the assumed values for the mean parameters in the design construction process.

Two types of integrals are used in the construction of a semi-Bayesian $D$-optimal mixed logit design. Each of them is usually approximated using large numbers of Monte Carlo draws. We call the draws used to compute the integral in (8) *prior draws* and label the draws used to approximate the integrals involved in each element in (4) *mixed logit draws*. Clearly, using large numbers of random draws for approximating these integrals leads to long computation times. In our algorithm, we therefore make systematic draws from the distributions $f(\boldsymbol{\beta})$ and $g(\boldsymbol{\mu}_{\boldsymbol{\beta}})$ rather than random ones. This allows us to approximate the integrals using much smaller numbers of draws, which reduces the computing time dramatically.

In total, the SB algorithm has four features that speed up the computation of semi-Bayesian $D$-optimal mixed logit designs: (i) the coordinate-exchange procedure, (ii) the small sample of *mixed logit draws*, (iii) the small sample of *prior draws*, and (iv) the fast update of the information matrix. The first three features are discussed in more detail below. The fast update of the information matrix, the fourth feature of the proposed algorithm, exploits the fact that the total information matrix in (4) is the sum of all per choice set information matrices. A change to one of the profiles in a choice set therefore requires only updating the information matrix corresponding to that choice set.

### 3.1. Coordinate-Exchange Algorithm

In this paper, we use the coordinate-exchange algorithm proposed by Meyer and Nachtsheim (1995) and introduced in the marketing literature by Kessels et al. (2008). Unlike the algorithm in Kessels et al. (2006), the coordinate-exchange algorithm does not require the construction of a candidate set of profiles. This property is especially important when the design involves a large number of attributes and attribute levels. Furthermore, it is also a computationally efficient algorithm. It replaces only one coordinate or attribute level of a profile at each step. Therefore, the coordinate-exchange algorithm is a special case of the profile-exchange algorithm.

For each run of the algorithm, a starting design with $K$ profiles in each of $S$ choice sets is constructed by randomly generating attribute levels for each of the $K \times S$ profiles in the design. Each attribute level in the starting design is then exchanged with all possible levels of that attribute. A level change is accepted if and only if it results in a better $D_{\mathrm{SBM}}$-error. The first iteration terminates when the algorithm has found the best exchange for all attributes of all profiles of the design. After that, the algorithm goes back to the first attribute of the first profile in the design and continues until no substantial improvement is possible any more. To avoid poorly local optima, we used 1,000 runs of the algorithm to find the designs reported in this paper.

### 3.2. Halton Sequences for the Mixed Logit Draws

The integrals involved in the information matrix (4) can be approximated using a sample of *mixed logit draws*. The well-known Monte Carlo simulation method is frequently applied in practice for this purpose. This method, however, requires large numbers of draws and has a slow asymptotic convergence rate (Bhat 2001). The generation and application of a small sample of intelligent draws from a distribution rather than a large sample of random ones has therefore been the subject of intensive research in recent years (Glasgow 2001, Sloan and Wozniakowski 1998, Train 2000). Especially Halton sequences, which produce uniformly distributed points over the domain of the integrals that have to be computed, have received a lot of attention. Compared with the random draws employed in the Monte Carlo method, a small sample of Halton draws converges faster and yields smaller simulation errors. Bhat (2001) applied Halton sequences for estimation of the mixed logit model and found that, with 125 Halton draws, the simulation error is half as large as with 1,000 random draws and even smaller than with 2,000 random draws. Encouraged by his findings, we use a small sample of Halton draws to approximate the integral in (4). Note that the number of Halton draws needed depends on the problem considered.

To determine how many Halton draws should be sufficient to generate efficient designs for our design settings, we considered three different numbers of Halton draws: 100, 200, and 250 draws and two different numbers of random Monte Carlo draws: 1,000 and 1,500. The design problem considered was the construction of a choice experiment with 12 choice sets of three alternatives described by four attributes. Each attribute had three levels. A short-hand notation for this setting is $3^4/3/12$. Five designs were then constructed by using the three different numbers of Halton draws and the two different numbers of random draws, respectively. To examine how well these different draws perform in constructing efficient mixed logit designs, we compared these five designs in terms of design efficiencies across 100 true parameter vectors. We denote the $r$th true parameter vector by $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$. Each $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$ was generated from the multivariate normal distribution with covariance matrix $\mathbf{I}_8$ and centered around $\boldsymbol{\mu}_0$, the mean parameter vector assumed when constructing the designs. The true heterogeneity parameter vector was assumed to be $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$. The idea behind this comparison is as follows: if any of these five different numbers of draws works well in generating an efficient mixed logit design, then it will lead to a design with a small $D_M$-error. For each $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$, the $D_M$-error value for each of these five designs was evaluated using 100,000 random draws to approximate the integrals in (4). Finally, we averaged all the $D_M$-error values over 100 true parameter values. These averages can be found in Table 1. It is clear that the design generated by 250 Halton draws, on average, had the best performance. This implies that 250 Halton draws enable us to find better designs than 1,000 and 1,500 Monte Carlo draws. The detailed construction procedures for Halton draws and the reason why these draws perform better than random draws can be found in Train (2000).

### 3.3. Small Number of Prior Draws

The integral that has to be computed for evaluating the $D_{\mathrm{SBM}}$-error in (8) does not require as many as 250 draws. Our own computational work as well as that described in Kessels et al. (2008) suggests that a designed sample of 20 *prior draws* instead of a Monte Carlo sample of 1,000 draws yields satisfactory results. The designed sample of 20 *prior draws* provides only a rough approximation of the integral in (8), but this turns out to be sufficient for the coordinate-exchange procedure to work well. It allows the execution of a large number of runs of the coordinate-exchange algorithm for searching the best design within a reasonable time. The methodology used by Kessels et al. (2008) to generate 20-point sets of *prior draws* performs slightly better in terms of design efficiency than Halton sequences with 20 draws, and it is effective in finding semi-Bayesian $D$-optimal designs. Note that this small designed sample cannot be used for the *mixed logit draws* because it leads to an approximation of the integrals involved in (4) that is too poor for the purpose of finding optimal designs for the mixed logit model.

The designed sample suggested by Kessels et al. (2008) starts from the fact that multivariate normal distributions, such as the ones used here as prior distributions for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, are symmetric around the prior mean. Therefore, each parameter on a $p$-dimensional hypersphere for a given radius has the same density. The key is to generate 20 prior parameter vectors that are uniformly distributed on such a sphere. On the sphere, the minimum distance to a neighboring vector from any of the vectors is the same for all of them. In this way, these 20 parameter vectors sample different directions away from the prior mean fairly. In Technical Appendix A, which can be found at http://mktsci.pubs.informs.org, we summarized the procedure for constructing the 20-point sample and we show in detail how this constructed sample is used when searching an efficient mixed logit design.

### 3.4. Comparing the SB Algorithm with a Benchmark Algorithm

To demonstrate how efficient the SB algorithm is relative to a benchmark algorithm in searching for a good design within a reasonable time, we report some computational results for a design setting $3^2/3/6$. The benchmark algorithm used random samples of 1,000 *prior draws* and 1,000 *mixed logit draws*, and it updates the information matrix for the whole sample for each change. For this small example, the SB algorithm is approximately 300 times faster than the benchmark algorithm for one run of the coordinate-exchange algorithm. The best designs found by these two algorithms have the same $D_{\mathrm{SBM}}$-error of 6.1774. Although the SB algorithm needs more runs to obtain the best design than the benchmark algorithm, its total computation time for the SB algorithm is far less: the benchmark algorithm took approximately 311 hours, whereas the SB algorithm needed less than 3 hours to find the best design. Further compelling evidence of the practical value of the SB algorithm is given in Technical Appendix B (found at http://mktsci.pubs.informs.org), which shows a comparison of the computational effectiveness of the SB algorithm and the benchmark algorithm by means of the estimated expected efficiencies of the designs generated by the two algorithms. Note that for the $3^4/3/12$ design problem, the leading example used in this paper, it is not feasible to obtain an efficient design using the benchmark algorithm because one single run of the algorithm requires more than 65 hours.

**Table 1** Comparing the Performance of Designs Generated with Different Numbers of Halton Draws and Monte Carlo Draws

| | Halton 100 | Halton 200 | Halton 250 | Monte Carlo 1,000 | Monte Carlo 1,500 |
|---|---|---|---|---|---|
| $D_M$-error | 5.61 | 5.56 | 5.25 | 5.58 | 5.44 |

## 4. Design Evaluation

In this section, we compare the performance of eight types of designs under different conditions. We investigate which designs allow for efficient parameter estimation and accurate predictions when there is heterogeneity among the respondents and when prior information about the mean and variance of the individual-level coefficients is incorrect. We are interested in quantifying the benefits of using semi-Bayesian mixed logit designs in particular. The results reported here are for a design problem with specification $3^4/3/12$.

### 4.1. Set of Design Options

An overview of the eight designs used in the comparison study is given in Table 2. In the table, $\mathbf{1}_8$ and $\mathbf{0}_8$ denote an $8 \times 1$ vector of ones and an $8 \times 1$ vector of zeros, respectively. The first three designs in our study are semi-Bayesian $D$-optimal designs for the mixed logit model. Each of these designs was constructed using the same prior distribution $g(\boldsymbol{\mu}_{\boldsymbol{\beta}})$ for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ but with a different prior value for the heterogeneity vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$. The prior distribution $g(\boldsymbol{\mu}_{\boldsymbol{\beta}})$ was multivariate normal with mean $\boldsymbol{\mu}_0 = [-0.5, 0, -0.5, 0, -0.5, 0, -0.5, 0]'$ and covariance matrix $\mathbf{I}_8$, with $\mathbf{I}_8$ the eight-dimensional identity matrix. The elements of the heterogeneity vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ were 1.5, 1, and 0.5 for the three designs. We denote them as $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 1.5[\mathbf{1}_8]$, $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$, and $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 0.5[\mathbf{1}_8]$, respectively. These three designs enable us to quantify the effect of misspecifying the heterogeneity parameters in the design construction. The fourth design in our study is a locally $D$-optimal design for the mixed logit model. This design ignores the prior parameter uncertainty. It was constructed using the same heterogeneity vector, $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$, as the second semi-Bayesian $D$-optimal mixed logit design. Comparing these two designs allows us to quantify the advantage of the semi-Bayesian approach over the locally optimal design approach.

Furthermore, two Bayesian $D$-optimal designs for a multinomial logit model, which assumes respondent homogeneity (i.e., $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{0}_8$), were used in the study. These two designs were computed using the prior mean vector $\boldsymbol{\mu}_0$ and variance–covariance matrices $9\mathbf{I}_8$ and $\mathbf{I}_8$. This enables us to investigate whether a larger prior uncertainty helps to overcome the fact that optimal designs for the multinomial logit model are obtained ignoring the respondent heterogeneity. In addition, we also used a locally $D$-optimal design for the multinomial logit model in our study because it is a design that has received considerable attention in the marketing literature (Bunch et al. 1996, Huber and Zwerina 1996, Zwerina et al. 1996). This design enables us to examine the loss in estimation efficiency and in predictive accuracy that results from ignoring the respondents' heterogeneity and the parameter uncertainty when constructing conjoint choice designs.

The final design in our study is a nearly orthogonal design generated using the Sawtooth software. Such designs are commonly used by researchers who do not have access to algorithms for computing optimal designs for multinomial logit models or for mixed logit models.

**Table 2** Overview of the Designs Used in the Simulation Study

| Design type | Mean prior | Heterogeneity prior |
|---|---|---|
| 1. Semi-Bayesian $D$-optimal, mixed logit | $\boldsymbol{\mu}_{\beta} \sim N(\boldsymbol{\mu}_0, \mathbf{I}_8)$ | $\boldsymbol{\sigma}_{\beta} = 1.5[\mathbf{1}_8]$ |
| 2. Semi-Bayesian $D$-optimal, mixed logit | $\boldsymbol{\mu}_{\beta} \sim N(\boldsymbol{\mu}_0, \mathbf{I}_8)$ | $\boldsymbol{\sigma}_{\beta} = \mathbf{1}_8$ |
| 3. Semi-Bayesian $D$-optimal, mixed logit | $\boldsymbol{\mu}_{\beta} \sim N(\boldsymbol{\mu}_0, \mathbf{I}_8)$ | $\boldsymbol{\sigma}_{\beta} = 0.5[\mathbf{1}_8]$ |
| 4. Locally $D$-optimal, mixed logit | $\boldsymbol{\mu}_{\beta} = \boldsymbol{\mu}_0$ | $\boldsymbol{\sigma}_{\beta} = \mathbf{1}_8$ |
| 5. Bayesian $D$-optimal, multinomial logit | $\boldsymbol{\mu}_{\beta} \sim N(\boldsymbol{\mu}_0, 9\mathbf{I}_8)$ | $\boldsymbol{\sigma}_{\beta} = \mathbf{0}_8$ |
| 6. Bayesian $D$-optimal, multinomial logit | $\boldsymbol{\mu}_{\beta} \sim N(\boldsymbol{\mu}_0, \mathbf{I}_8)$ | $\boldsymbol{\sigma}_{\beta} = \mathbf{0}_8$ |
| 7. Locally $D$-optimal, multinomial logit | $\boldsymbol{\mu}_{\beta} = \boldsymbol{\mu}_0$ | $\boldsymbol{\sigma}_{\beta} = \mathbf{0}_8$ |
| 8. Nearly orthogonal | — | — |

### 4.2. Specification of the True Data-Generating Process

In our simulation study, we evaluated the eight designs in Table 2 under five conditions. In each case, a different true parameter set $\boldsymbol{\Omega}_i$ was used to draw 1,000 values of the true mean vector $\boldsymbol{\mu}_{\boldsymbol{\beta}}$. The heterogeneity vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ was fixed at $\mathbf{1}_8$ for each condition.

For the true parameter sets $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, and $\boldsymbol{\Omega}_3$, the true mean parameter vector $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, used for generating the data, was drawn from distributions centered around $\boldsymbol{\mu}_0$ with covariance matrices $0.25\mathbf{I}_8$, $\mathbf{I}_8$, and $2.25\mathbf{I}_8$, respectively. As a consequence, the true parameter sets $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, and $\boldsymbol{\Omega}_3$ are characterized by an increasing degree of mean parameter misspecification. For true parameter sets $\boldsymbol{\Omega}_4$ and $\boldsymbol{\Omega}_5$, the true mean $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ was drawn from distributions with a mean that is different from $\boldsymbol{\mu}_0$ and with covariance matrix $\mathbf{I}_8$. For true parameter set $\boldsymbol{\Omega}_4$, the mean vector was $-\boldsymbol{\mu}_0$, whereas for true parameter set $\boldsymbol{\Omega}_5$, it was $-\boldsymbol{\mu}_0 + 0.5[\mathbf{1}_8]$. The degree of mean vector misspecification is therefore even larger for these sets of true parameters than for $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, and $\boldsymbol{\Omega}_3$. An overview of the five true parameter sets is given in Table 3.

### 4.3. Performance Criteria

In our simulation study, we have used four different criteria to evaluate the eight different design options.

**Table 3** True Parameter Sets' Specifications

| | |
|---|---|
| $\boldsymbol{\Omega}_1$ | $\boldsymbol{\mu}_{\beta} \sim N(\boldsymbol{\mu}_0, 0.25\mathbf{I}_8)$ |
| $\boldsymbol{\Omega}_2$ | $\boldsymbol{\mu}_{\beta} \sim N(\boldsymbol{\mu}_0, \mathbf{I}_8)$ |
| $\boldsymbol{\Omega}_3$ | $\boldsymbol{\mu}_{\beta} \sim N(\boldsymbol{\mu}_0, 2.25\mathbf{I}_8)$ |
| $\boldsymbol{\Omega}_4$ | $\boldsymbol{\mu}_{\beta} \sim N(-\boldsymbol{\mu}_0, \mathbf{I}_8)$ |
| $\boldsymbol{\Omega}_5$ | $\boldsymbol{\mu}_{\beta} \sim N(-\boldsymbol{\mu}_0 + 0.5[\mathbf{1}_8], \mathbf{I}_8)$ |

The first two criteria, defined in §4.3.1, quantify the precision of the model estimation. The other two criteria, defined in §4.3.2, measure the accuracy of the predictions.

**4.3.1. Relative Local *D*-Efficiency.** First, we used the relative local *D*-efficiency (RLD) proposed by Woods et al. (2006) as a measure for evaluating the performance of the eight designs in terms of the precision of the model estimation. For that purpose, we constructed a locally *D*-optimal design for the mixed logit model for each draw of the true mean parameter $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$, while fixing the heterogeneity vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ at $\mathbf{1}_8$. We denote that locally optimal design by $\mathbf{X}_{\boldsymbol{\mu}_{\boldsymbol{\beta}}^r}$. For a given $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$, the RLD-efficiency of a specific design $\mathbf{X}$ is then computed by comparing its $D_M$-error to that of the locally *D*-optimal mixed logit design $\mathbf{X}_{\boldsymbol{\mu}_{\boldsymbol{\beta}}^r}$ assuming that $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{\boldsymbol{\beta}}^r$:

$$\mathrm{RLD}(\mathbf{X}, \boldsymbol{\mu}_{\boldsymbol{\beta}}^r) = \frac{D_M\text{-error}(\mathbf{X}_{\boldsymbol{\mu}_{\boldsymbol{\beta}}^r} \mid \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{\boldsymbol{\beta}}^r)}{D_M\text{-error}(\mathbf{X} \mid \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{\boldsymbol{\beta}}^r)}. \tag{10}$$

This expression lies between zero and one. RLD-efficiencies close to one indicate that the design $\mathbf{X}$ provides nearly as much information on the unknown model parameters as the locally *D*-optimal design when $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{\boldsymbol{\beta}}^r$. Designs that have RLD-efficiencies close to one for all draws $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$ from a certain true parameter set $\boldsymbol{\Omega}_i$ are desirable.

In our simulation study, we computed RLD-efficiencies for every draw $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$ for each of the five true parameter sets. Histograms that visualize the distribution of the resulting 1,000 RLD-efficiencies for each of the eight design options in Table 2 for true parameter set $\boldsymbol{\Omega}_2$ are shown in Figure 1. A summary of the results for all five true parameter sets can be found in Table 4. These results are discussed in detail in §5.1.

In addition to the RLD-efficiencies, we also computed the percent reduction in the number of observations that a design, $\mathbf{X}_1$, requires to produce the same expected $D_M$-error as another design, $\mathbf{X}_2$. This measure assesses the relative performance of a pair of designs and is defined by

$$1 - \frac{D_M\text{-error}(\mathbf{X}_1 \mid \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{\boldsymbol{\beta}}^r)}{D_M\text{-error}(\mathbf{X}_2 \mid \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{\boldsymbol{\beta}}^r)} \tag{11}$$

for a given draw $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$. We averaged these values over all draws of a particular true parameter set to quantify the extent to which the former design, $\mathbf{X}_1$, performs better than $\mathbf{X}_2$, on average.

**4.3.2. Expected Root Mean-Squared Prediction Error.** To assess the predictive accuracy of the constructed designs, we computed the expected root mean-squared prediction error, $\mathrm{ERMSE}_P(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}})$, for a given $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$. The $\mathrm{ERMSE}_P(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}})$ compares the predicted probabilities to the true probabilities based on a $3^4/2/12$ holdout design generated using Sawtooth and is defined as

$$\begin{aligned}\mathrm{ERMSE}_P&(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}})\\ &= \int [(\pi(\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{\beta}}) - \pi(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}}))'(\pi(\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{\beta}}) - \pi(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}}))]^{1/2}\\ &\quad \cdot \phi(\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{\beta}})\, d\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}}\, d\hat{\boldsymbol{\sigma}}_{\boldsymbol{\beta}},\end{aligned} \tag{12}$$

where $\pi(\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{\beta}})$ is the vector containing the predicted probabilities computed using the parameter estimates $\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\sigma}}_{\boldsymbol{\beta}}$, $\pi(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}})$ is the vector of the probabilities obtained using the true parameter vectors $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$, and $\phi(\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{\beta}})$ is the asymptotic distribution of the parameter estimates. For a design $\mathbf{X}$ and true parameter vectors $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$, the asymptotic distribution $\phi(\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{\beta}})$ of the parameter estimates is the multivariate normal distribution with mean vector

**Table 4** The Mean, the Standard Deviation, the Minimum and the Maximum of the RLD-Efficiencies of the Eight Designs

| Designs | $\boldsymbol{\Omega}_1$ | $\boldsymbol{\Omega}_2$ | $\boldsymbol{\Omega}_3$ | $\boldsymbol{\Omega}_4$ | $\boldsymbol{\Omega}_5$ |
|---|---|---|---|---|---|
| Semi-Bayesian mixed logit $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 1.5[\mathbf{1}_8]$ | 0.824 (0.068) [0.526, 0.988] | 0.663 (0.095) [0.286, 0.994] | 0.543 (0.111) [0.163, 0.974] | 0.566 (0.090) [0.274, 0.873] | 0.487 (0.078) [0.199, 0.812] |
| Semi-Bayesian mixed logit $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$ | 0.841 (0.057) [0.571, 0.964] | 0.688 (0.091) [0.333, 0.966] | 0.560 (0.113) [0.141, 0.947] | 0.572 (0.093) [0.278, 0.999] | 0.473 (0.088) [0.203, 0.828] |
| Semi-Bayesian mixed logit $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 0.5[\mathbf{1}_8]$ | 0.707 (0.039) [0.506, 0.792] | 0.624 (0.069) [0.317, 0.784] | 0.522 (0.099) [0.152, 0.748] | 0.540 (0.070) [0.273, 0.845] | 0.462 (0.072) [0.236, 0.711] |
| Locally *D*-optimal mixed logit $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$ | 0.818 (0.080) [0.509, 0.996] | 0.636 (0.104) [0.257, 0.992] | 0.510 (0.116) [0.083, 0.997] | 0.521 (0.094) [0.233, 0.884] | 0.435 (0.088) [0.141, 0.801] |
| Bayesian multinomial with covariance $9\mathbf{I}_8$ | 0.601 (0.037) [0.462, 0.696] | 0.542 (0.056) [0.282, 0.683] | 0.465 (0.082) [0.133, 0.675] | 0.486 (0.060) [0.277, 0.864] | 0.435 (0.066) [0.221, 0.656] |
| Bayesian multinomial with covariance $\mathbf{I}_8$ | 0.514 (0.029) [0.397, 0.598] | 0.474 (0.044) [0.225, 0.588] | 0.410 (0.069) [0.070, 0.568] | 0.463 (0.049) [0.266, 0.859] | 0.436 (0.042) [0.211, 0.559] |
| Locally *D*-optimal multinomial logit | 0.279 (0.034) [0.174, 0.381] | 0.319 (0.033) [0.080, 0.401] | 0.299 (0.051) [0.009, 0.423] | 0.322 (0.037) [0.181, 0.611] | 0.333 (0.033) [0.110, 0.430] |
| Nearly orthogonal design | 0.288 (0.032) [0.173, 0.369] | 0.312 (0.037) [0.123, 0.415] | 0.286 (0.059) [0.020, 0.407] | 0.322 (0.039) [0.184, 0.576] | 0.310 (0.044) [0.115, 0.453] |

$(\boldsymbol{\mu}_{\boldsymbol{\beta}}', \boldsymbol{\sigma}_{\boldsymbol{\beta}}')'$ and covariance matrix $\mathcal{I}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}} \mid \mathbf{X})^{-1}$. We approximated the integral in (12) by using 1,000 random draws from that asymptotic distribution. Note that the ERMSE$_P$ value depends on the values of true parameters $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$. We computed ERMSE$_P$ values for 1,000 draws of $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ from each of the five true parameter sets given in Table 3 for the eight designs in Table 2. Histograms of the ERMSE$_P$ values for true parameter set $\boldsymbol{\Omega}_2$ are shown in Figure 2 and discussed in §5.2, along with the results for the other true parameter sets.

In addition, to quantify the extent to which a design, $\mathbf{X}_1$, allows for better predictions than another design, $\mathbf{X}_2$, we computed the percent decrease in prediction error by using the former design instead of the latter. This measure is defined by

$$1 - \frac{\text{ERMSE}_P(\mathbf{X}_1 \mid \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{\boldsymbol{\beta}}^r)}{\text{ERMSE}_P(\mathbf{X}_2 \mid \boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{\boldsymbol{\beta}}^r)} \tag{13}$$

for a given draw $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$. The values we report are averages over the total number of draws.

## 5. Simulation Results

### 5.1. Efficiency of Parameter Estimation

In this section, we evaluate the performance of the eight designs in Table 2 in terms of the efficiency of parameter estimation when the true mean $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ is taken from the five true parameter sets in Table 3. The aim is to find out which design option is more robust against misspecification of the mean and the variance of the heterogeneity distribution. Table 4 presents the mean, the standard deviation, the minimum, and the maximum values of the RLD-efficiencies of the eight designs for all five true parameter sets. Note that the minimum and the maximum of the RLD-efficiencies are shown between brackets on the second row for each design option. These values indicate the worst and best relative local efficiencies provided by the different designs under each true parameter set. A design with a minimum RLD-efficiency value close to zero means that, for some true mean parameter values, almost no information can be provided by that design. RLD-efficiency values close to one indicate that the design is almost as good as the locally optimal design. Four common features occur across the five true parameter sets in Table 4.

First, all of the optimal mixed logit designs perform substantially better than the optimal designs that were constructed ignoring the respondent heterogeneity. This holds in particular for true parameter set $\boldsymbol{\Omega}_1$, where the mean parameter vectors $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$ are all close to the one assumed when constructing the optimal designs, $\boldsymbol{\mu}_0$. Furthermore, the semi-Bayesian mixed logit designs constructed with heterogeneity parameter vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$ and with overspecified heterogeneity vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 1.5[\mathbf{1}_8]$ provide the best estimation accuracies. This implies that, with these two designs, fewer observations are needed to achieve the same precision for the parameter estimates as with the other designs.

Second, the comparison of the semi-Bayesian $D$-optimal designs for the mixed logit model constructed with $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 1.5[\mathbf{1}_8]$ and with $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 0.5[\mathbf{1}_8]$ to the one obtained using $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$ shows that the loss in RLD-efficiencies for underspecifying $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ is larger than that for overspecifying $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ to the same extent. However, the design constructed with underspecified $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ is still more efficient than the designs constructed for the multinomial logit model, which ignores the respondent heterogeneity. In addition, the small differences in efficiency between the first and the second designs in Table 4 indicate that overspecifying the heterogeneity vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ does not have a large negative impact on the efficiency of parameter estimation. To see whether specifying a too large value for $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ will lead to a design that performs much worse than the design constructed with the correctly specified $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$, we considered another three semi-Bayesian mixed logit designs, which were constructed with heterogeneity parameters $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 2[\mathbf{1}_8]$, $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 3[\mathbf{1}_8]$, and $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 4[\mathbf{1}_8]$, respectively. Table 5 presents the percent reductions in the number of observations by using $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$ instead of overspecifying the heterogeneity parameters. Note that the larger the values in Table 5, the larger the extent to which the design with correctly specified $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ outperforms those with overspecified $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$. It seems that the semi-Bayesian mixed logit design generally allows a large range of overspecification of the heterogeneity parameters. For instance, it is seen that specifying $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 2[\mathbf{1}_8]$, which is two times the true value, still leads to a design that is only slightly less efficient in parameter estimation than the design with correctly specified $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ for the first four true parameter sets. For $\boldsymbol{\Omega}_5$, where the misspecification of the mean parameter vector is considerably larger than for the other true parameter sets, the negative values demonstrate that the semi-Bayesian mixed logit design constructed by assuming a large heterogeneity vector

**Table 5 Percent Reduction in the Number of Observations by Using the Semi-Bayesian Design Constructed with the Correctly Specified $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ Relative to the Designs Obtained with Overspecified $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$**

| True parameter set | $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 1.5[\mathbf{1}_8]$ (%) | $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 2[\mathbf{1}_8]$ (%) | $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 3[\mathbf{1}_8]$ (%) | $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 4[\mathbf{1}_8]$ (%) |
|---|---|---|---|---|
| $\boldsymbol{\Omega}_1$ | 2.01 | 3.3 | 7.16 | 11.7 |
| $\boldsymbol{\Omega}_2$ | 3.4 | 4.22 | 5.92 | 11.33 |
| $\boldsymbol{\Omega}_3$ | 2.4 | 3.67 | 4.02 | 10.5 |
| $\boldsymbol{\Omega}_4$ | 0.49 | 2.57 | 1.39 | 6.00 |
| $\boldsymbol{\Omega}_5$ | −3.8 | −1.2 | −2.97 | 0.24 |

**Figure 1** **RLD-Efficiencies for Parameter Set $\Omega_2$**

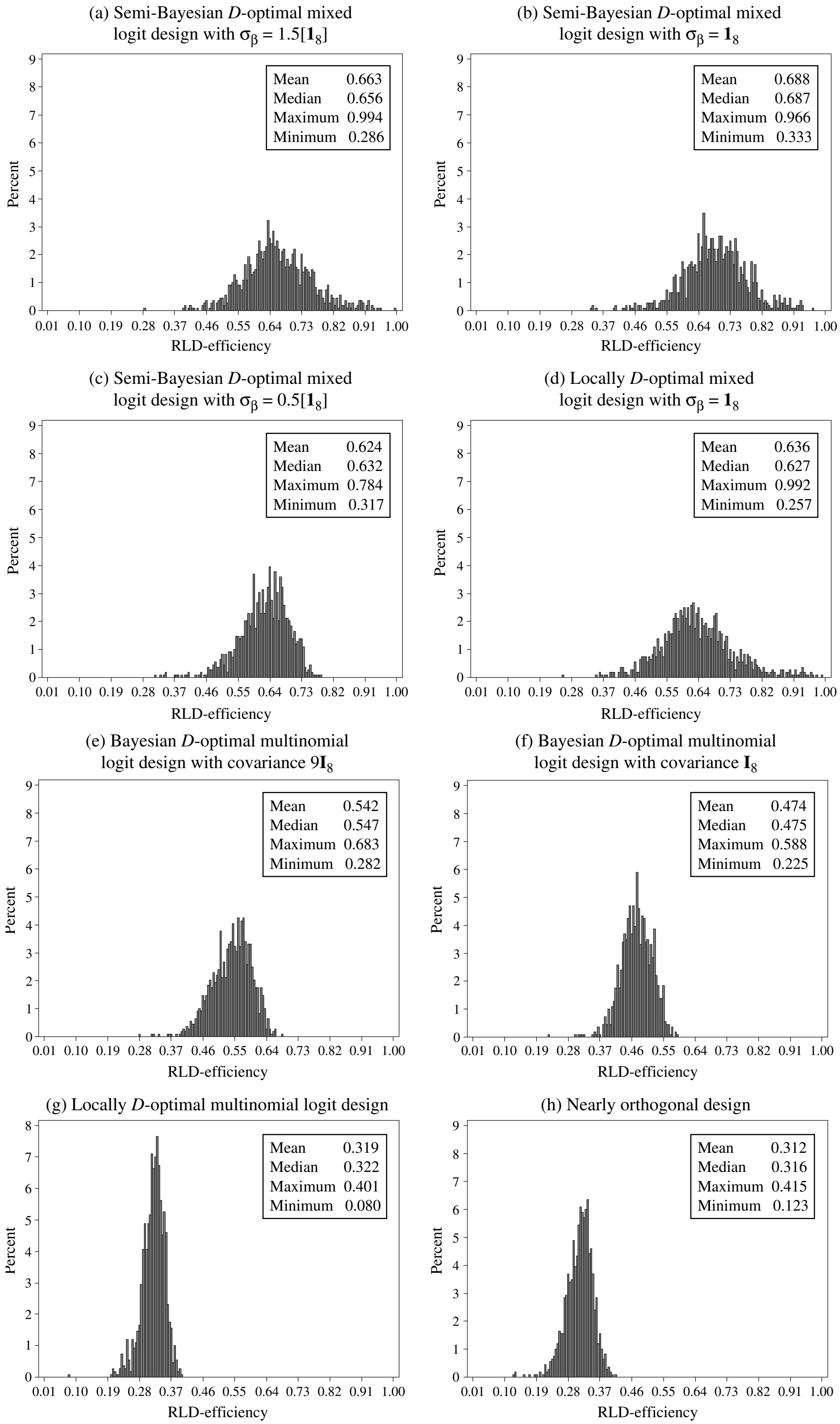
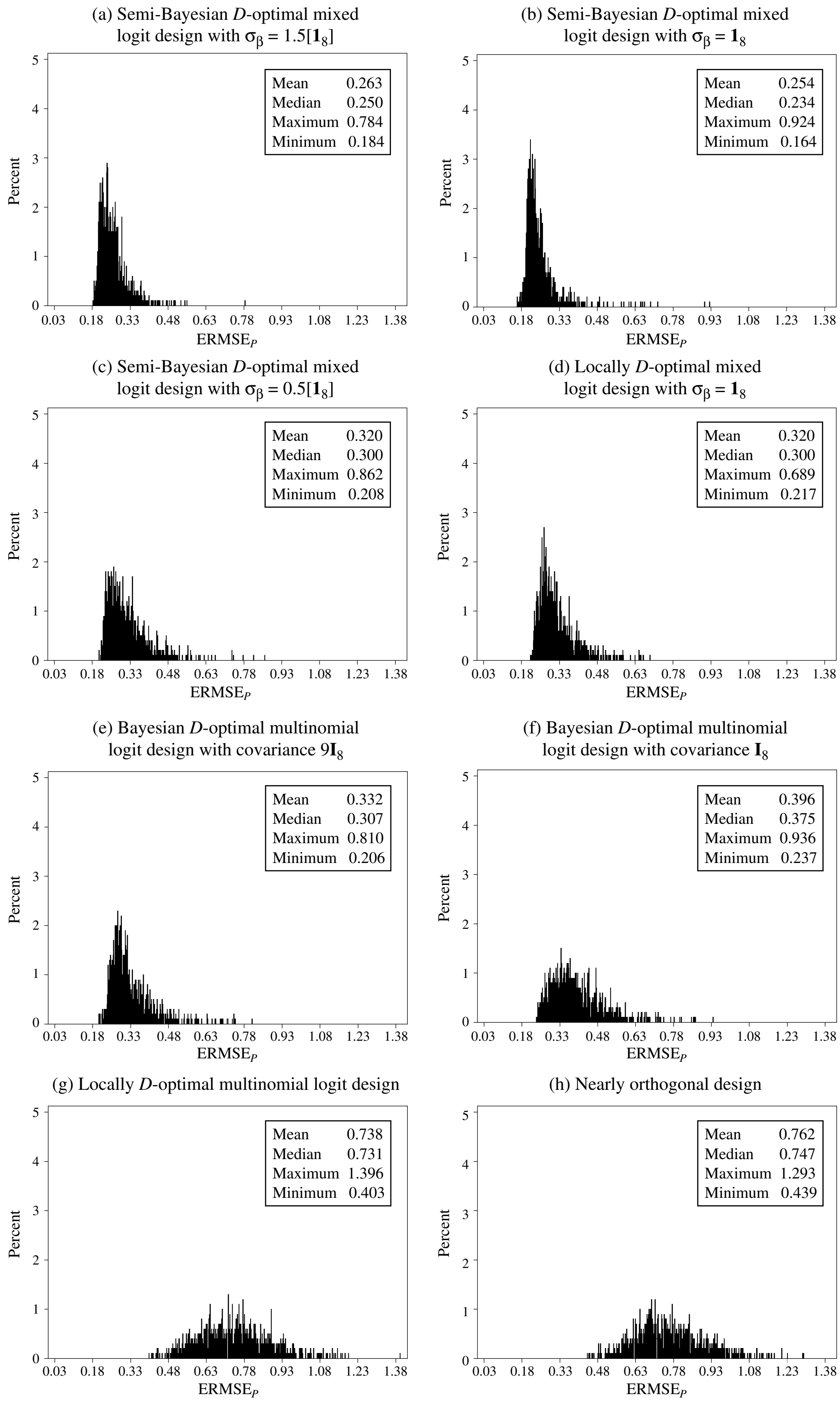


**Figure 2 Expected Root Mean-Squared Prediction Errors for Parameter Set $\Omega_2$**



is more robust to the misspecification of the mean parameters than when assuming a small heterogeneity. This result is useful when prior information about the mean parameter vector $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ is lacking. Sándor and Wedel (2002) obtained similar results in the context of locally *D*-optimal mixed logit designs. They showed that large heterogeneity parameters assumed in the design construction may help to account for the misspecification of the mean parameters.

Third, comparing the second to the fourth design in Table 4 demonstrates the benefit of using the semi-Bayesian approach. For each true parameter set, it turns out that the semi-Bayesian design outperforms the locally optimal design. On average, the semi-Bayesian mixed logit design requires 3.3%, 9.1%, and 11.5% fewer observations to have the same efficiency as the locally *D*-optimal mixed logit design for true parameter sets $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, and $\boldsymbol{\Omega}_3$, respectively. For true parameter sets $\boldsymbol{\Omega}_4$ and $\boldsymbol{\Omega}_5$, which are centered away from the true mean parameter vector $\boldsymbol{\mu}_0$ used during the design construction, the average reductions are 10% and 9.8%, respectively.

Finally, the locally *D*-optimal design for the multinomial logit model and the nearly orthogonal design yield substantially poorer RLD-efficiencies than the other design options under all conditions. This can be seen from Figure 1 by comparing panels (g) and (h) with the other panels: almost all RLD-efficiencies for the locally *D*-optimal multinomial logit design and for the nearly orthogonal design are markedly smaller than those for the other designs. The maximum RLD-efficiencies of these two designs are only slightly larger than the minimum RLD-efficiency of the semi-Bayesian mixed logit design constructed with $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$. In addition, for a true parameter set that accommodates a wide variety of mean parameters such as $\boldsymbol{\Omega}_3$, the minimum RLD-efficiency is 0.009 for the locally *D*-optimal multinomial logit design and 0.020 for the nearly orthogonal design. This indicates that, for some parameter values, the locally *D*-optimal multinomial logit design and the nearly orthogonal design yield almost no information. Another striking result is that, for all of the parameter sets, the average reductions in the number of observations required for the best semi-Bayesian *D*-optimal mixed logit design to obtain the same efficiency as the locally *D*-optimal multinomial logit design and the nearly orthogonal design are substantial. For instance, for parameter set $\boldsymbol{\Omega}_1$, the reductions are 66.6% and 65.4%, respectively. This shows that using locally *D*-optimal designs for the multinomial logit model or nearly orthogonal designs in the presence of respondent heterogeneity is not at all a good idea.

To investigate whether the results reported in this paper depend on the example that was chosen (four attributes acting at three levels), we conducted similar studies involving different numbers of attributes and unequal numbers of attribute levels, such as designs of the type $2 \times 3 \times 4/3/12$, and designs of the type $2 \times 2 \times 3/3/12$. These designs have the same number of profiles as the one used in the example described above, but the number of parameters involved is smaller. We found that the differences in the precision of the parameter estimates between the alternative designs become slightly smaller. For example, for parameter set $\boldsymbol{\Omega}_1$, the advantage of using the semi-Bayesian mixed logit design instead of the locally *D*-optimal multinomial logit design drops from 66.6% to 55% and to 51% for designs of type $2 \times 3 \times 4/3/12$ and of type $2 \times 2 \times 3/3/12$, respectively. In general, however, these studies lead to results similar to those discussed before.

### 5.2. Predictive Accuracy

In this section, we evaluate the eight designs in Table 2 in terms of their predictive accuracy. The purpose is to examine whether the semi-Bayesian *D*-optimal mixed logit designs, which are more efficient than the other designs in terms of precision of the parameter estimation, also lead to more accurate predictions. In this section, we focus on the results for true parameter set $\boldsymbol{\Omega}_2$. The results for the other four true parameter sets are similar, and they are only briefly discussed at the end of this section. Figure 2 shows the $\text{ERMSE}_P$ values of the eight designs for true parameter set $\boldsymbol{\Omega}_2$. Note that small $\text{ERMSE}_P$ values are desirable.

As can be seen in Figure 2, the semi-Bayesian *D*-optimal mixed logit design with $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$, the $\text{ERMSE}_P$ values for which are shown in panel (b), generally performs best in terms of predictive accuracy. Compared to the locally *D*-optimal mixed logit design with $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$ (see panel (d)), adopting a semi-Bayesian approach in the design construction, on average, leads to an 18.7% decrease in prediction error. The semi-Bayesian approach is thus substantially better than the locally optimal design approach when it comes to predicting choice probabilities in the presence of respondent heterogeneity.

The effects of misspecifying the heterogeneity vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ on the predictive accuracy can be examined by comparing the semi-Bayesian *D*-optimal mixed logit designs for $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 1.5[\mathbf{1}_8]$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 0.5[\mathbf{1}_8]$ to that for $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$. As can be seen from the panels (a)–(c) in Figure 2, the effects are similar to those for the RLD-efficiencies in the previous section. The design constructed using the correctly specified heterogeneity vector $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ produces slightly smaller $\text{ERMSE}_P$ values on average than the design with overspecified $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$. The average decrease in prediction error is only 2.66%. Underspecifying $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ leads to a larger loss in predictive accuracy than overspecifying it. Therefore, we can conclude that it is best to use large values for $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ to construct designs if there is uncertainty about these parameters. Note that the differences in predictive accuracy between the design with correctly specified $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$ and those with overspecified $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 2[\mathbf{1}_8]$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 3[\mathbf{1}_8]$ are also rather small. They amount to 3.64% and 6.57%, respectively.

By looking at panels (g) and (h) of Figure 2, it can be verified that the locally $D$-optimal design for the multinomial logit model and the nearly orthogonal design exhibit a large variation in $\text{ERMSE}_P$ values and lead to large $\text{ERMSE}_P$ values on average. Therefore, they can produce extremely inaccurate predictions. Compared to these two designs, the best semi-Bayesian $D$-optimal mixed logit design, the results for which are shown in panel (b), yields a 64.8% and a 65.9% average decrease in prediction error.

A comparison between the $\text{ERMSE}_P$ values in panel (e) for the Bayesian $D$-optimal design with covariance matrix $9\mathbf{I}_8$ and those in panel (f) for the Bayesian $D$-optimal design with covariance matrix $\mathbf{I}_8$ reveals that, when consumer heterogeneity is present, the Bayesian multinomial logit design produces better predictions when its prior distribution has a large variance. The decrease in prediction error is, on average, 12.5% if the former design is used instead of the latter. This shows that assuming a large variance about the prior parameters when constructing Bayesian multinomial logit designs is a way to offset some of the loss in prediction accuracy caused by ignoring the respondent heterogeneity. To see whether specifying a covariance matrix larger than $9\mathbf{I}_8$ further improves the predictive performance, we considered another three Bayesian multinomial logit designs constructed with prior covariance matrices $16\mathbf{I}_8$, $25\mathbf{I}_8$, and $36\mathbf{I}_8$. The results show that the predictive performance starts to decrease as soon as the covariance matrix approaches $25\mathbf{I}_8$.

Evidently, we performed similar studies for the other true parameter sets. The results are similar to those for $\boldsymbol{\Omega}_2$ except that, for true parameter sets $\boldsymbol{\Omega}_4$ and $\boldsymbol{\Omega}_5$, the semi-Bayesian $D$-optimal mixed logit design with $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 1.5[\mathbf{1}_8]$ has the best predictive performance. As indicated in Table 3, $\boldsymbol{\Omega}_4$ and $\boldsymbol{\Omega}_5$ are centered away from the mean parameter vector $\boldsymbol{\mu}_0$ that was utilized for the construction of the optimal designs. Therefore, the degree of misspecification of the mean parameters is quite high for these two parameter sets. The best predictive performance by the semi-Bayesian $D$-optimal mixed logit design with $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = 1.5[\mathbf{1}_8]$ implies that designs constructed with large heterogeneity parameters tend to be more robust against misspecification of the mean parameters in terms of the precision of prediction. Therefore, we suggest large values for specifying the heterogeneity vector in the design construction algorithm when sufficient information on the true mean parameter values is lacking. Note that, for these two true parameter sets, the semi-Bayesian mixed logit design constructed with $\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \mathbf{1}_8$ still outperforms the locally $D$-optimal mixed logit design and the designs that ignore the heterogeneity.

## 6. Semi-Bayesian vs. Fully Bayesian Design

As mentioned earlier, the fully Bayesian mixed logit design is hard to compute for a realistic problem. Therefore, we proposed the semi-Bayesian mixed logit design, which is less computationally demanding. In previous sections, we showed that the semi-Bayesian mixed logit design, in general, is quite robust to the misspecification of the mean parameters and heterogeneity parameters. Therefore, we believe that a semi-Bayesian mixed logit design would not perform significantly worse than the fully Bayesian mixed logit design in most cases. To demonstrate this, we conducted another study where we examined the relative performance of the semi-Bayesian mixed logit design and the fully Bayesian mixed logit design. The fully Bayesian design was constructed by minimizing the $D_{\text{BM}}$-error in (9). We assumed that $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ are independent such that their joint distribution $q(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}})$ can be written as the product of the prior distributions for $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$, that is, $g(\boldsymbol{\mu}_{\boldsymbol{\beta}})m(\boldsymbol{\sigma}_{\boldsymbol{\beta}})$. We used $g(\boldsymbol{\mu}_{\boldsymbol{\beta}}) \sim N(\boldsymbol{\mu}_0, \mathbf{I}_8)$ as the prior distribution for the mean parameters and assumed the elements $\sigma_i$ of $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ to be independent of each other. For each $\sigma_i$, we used an inverse gamma prior distribution with parameters 3 and 2, and thus with mean and variance both equal to one. The construction of the fully Bayesian mixed logit design is described in detail in the appendix.

We evaluate the performance of the fully Bayesian and the semi-Bayesian designs under 10 conditions. The true values of the mean parameters are drawn from each of the true parameter sets $\boldsymbol{\Omega}_i$ specified in Table 3. The heterogeneity parameters are drawn from two true parameter sets $\boldsymbol{\Gamma}_i$, $i = 1, 2$. $\boldsymbol{\Gamma}_1$ is defined by an inverse gamma distribution with parameters 3 and 2, and thus with mean and variance both equal to one. This is the prior distribution for each component of $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ used to construct the fully Bayesian design. $\boldsymbol{\Gamma}_2$ is defined by an inverse gamma distribution with parameters 2.5 and 1.5 and thus with mean equal to one and variance equal to two. As a consequence, the parameter set $\boldsymbol{\Gamma}_2$ contains more heterogeneous values for $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ than $\boldsymbol{\Gamma}_1$. For each condition, we drew 100 true mean parameters $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$ and 100 true heterogeneity parameters $\boldsymbol{\sigma}_{\boldsymbol{\beta}}^r$, leading to $100 \times 100 = 10{,}000$ different combinations of true parameter values for $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}^r$. For each combination of $\boldsymbol{\mu}_{\boldsymbol{\beta}}^r$ and $\boldsymbol{\sigma}_{\boldsymbol{\beta}}^r$, we

**Table 6** The Percentage Improvement by Using the Fully Bayesian Approach Instead of the Semi-Bayesian Approach in Terms of Estimation Efficiency and Predictive Accuracy

| Parameter space for $\boldsymbol{\mu}_\beta$ | Parameter space for $\boldsymbol{\sigma}_\beta$ | Improvement in estimation efficiency (%) | Improvement in predictive accuracy (%) |
|---|---|---|---|
| $\boldsymbol{\Omega}_1$ | $\boldsymbol{\Gamma}_1$ | −1.4 | −4.5 |
| | $\boldsymbol{\Gamma}_2$ | −0.7 | −5.87 |
| $\boldsymbol{\Omega}_2$ | $\boldsymbol{\Gamma}_1$ | 1.2 | −2.4 |
| | $\boldsymbol{\Gamma}_2$ | 1.67 | −1.17 |
| $\boldsymbol{\Omega}_3$ | $\boldsymbol{\Gamma}_1$ | 2.6 | 4.1 |
| | $\boldsymbol{\Gamma}_2$ | 3.07 | 4.5 |
| $\boldsymbol{\Omega}_4$ | $\boldsymbol{\Gamma}_1$ | 2.2 | 1.05 |
| | $\boldsymbol{\Gamma}_2$ | 2.63 | 1.38 |
| $\boldsymbol{\Omega}_5$ | $\boldsymbol{\Gamma}_1$ | 6.47 | 5.60 |
| | $\boldsymbol{\Gamma}_2$ | 6.80 | 6.41 |

calculated the percentage of improvement by using the fully Bayesian instead of the semi-Bayesian mixed logit designs with respect to estimation efficiency and predictive accuracy. The average results are summarized in Table 6.

Table 6 shows that, given a true parameter set $\boldsymbol{\Omega}_i$ for $\boldsymbol{\mu}_\beta$, the improvement by using the fully Bayesian approach instead of the semi-Bayesian approach tends to be larger for $\boldsymbol{\Gamma}_2$, which corresponds to a larger degree of misspecification of the heterogeneity parameters than $\boldsymbol{\Gamma}_1$. The comparison across true parameter sets $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, and $\boldsymbol{\Omega}_3$ for a given true parameter set $\boldsymbol{\Gamma}_i$ shows that, as the misspecification of the mean parameter vector $\boldsymbol{\mu}_\beta$ increases, the fully Bayesian approach is more preferable. Compared to true parameter set $\boldsymbol{\Omega}_4$, the misspecification of the mean parameters in $\boldsymbol{\Omega}_5$ is more severe and hence the advantage of using a fully Bayesian approach is larger for $\boldsymbol{\Omega}_5$. Furthermore, it seems that misspecifying $\boldsymbol{\mu}_\beta$ has a larger impact on the relative performance of the fully and semi-Bayesian design than misspecifying $\boldsymbol{\sigma}_\beta$. From the results for $\boldsymbol{\Omega}_1$, it can be seen that the semi-Bayesian design tends to perform better when the misspecification of the mean parameter $\boldsymbol{\mu}_\beta$ is small no matter what the size of the heterogeneity parameters. Only when the prior information about the possible value of the mean parameters is totally wrong, it is important to use the fully Bayesian approach in constructing mixed logit designs. This is demonstrated by the results for parameter set $\boldsymbol{\Omega}_5$. Because we believe that entirely wrong prior information is unlikely to occur in practice, we can conclude that there seems to be little benefit from using a fully Bayesian approach.

In addition, we conducted a similar study that includes a fully Bayesian design constructed with heterogeneity parameters drawn from an inverse gamma prior distribution with parameters 2.5 and 1.5 and thus with mean equal to one and variance equal to two. We investigated its performance under the same true parameter sets as previously. The improvements by using the fully Bayesian instead of the semi-Bayesian mixed logit designs range from −3.07% to 6.81% in terms of parameter estimation and from −7.90% to 8.00% in terms of predictive accuracy. This supports the conclusion that the semi-Bayesian design is not too sensitive to the misspecification of the heterogeneity parameters. Because the fully Bayesian design cannot be computed within a reasonable time, it seems that using semi-Bayesian mixed logit designs is a sensible thing to do in practice.

## 7. Conclusions

In this paper, we propose a new algorithm for computing *D*-optimal conjoint choice designs for mixed logit models in the presence of respondent heterogeneity. We used the algorithm to generate semi-Bayesian and fully Bayesian *D*-optimal designs. The semi-Bayesian approach differs from the fully Bayesian one in that it uses the simplifying assumption that the extent to which respondents are heterogenous is known. Although this assumption is unrealistic, it is absolutely required to keep the computations doable for practical conjoint design problems. It turns out, however, that the semi-Bayesian approach does not entail dramatic losses in estimation and predictive accuracy in the instances where we did generate the fully Bayesian designs at the expense of a major computational effort. In fact, for several scenarios we looked at, the semi-Bayesian designs led to a better precision of estimation and/or a better predictive accuracy. Moreover, the only cases where the semi-Bayesian designs performed substantially worse than the fully Bayesian design corresponded to a scenario where all prior information available about the parameters in $\boldsymbol{\mu}_\beta$ was entirely wrong. For these reasons, the focus in this article is on the performance of the more practical semi-Bayesian mixed logit design. Eight designs ranging from a nearly orthogonal design generated using Sawtooth and a simple locally *D*-optimal design for the multinomial logit model to the complex semi-Bayesian *D*-optimal mixed logit designs are compared in terms of estimation efficiency and predictive accuracy. The simulation study leads to the conclusion that the semi-Bayesian *D*-optimal designs for the mixed logit model consistently perform better than the other designs. Using nearly orthogonal designs and locally *D*-optimal designs for a multinomial logit model, both of which ignore respondent heterogeneity and parameter uncertainty when constructing the choice design, leads to extremely poor estimates and predictions and clearly should be avoided whenever respondents are heterogeneous. Semi-Bayesian mixed logit designs constructed with large heterogeneity prior parameters are most robust against the misspecification of the values for the prior mean of the individual-level coefficients. Therefore, specifying large values for $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$ is the best strategy when one has no idea about the possible values of the mean of the individual-level coefficients before conducting the survey. An interesting finding is that assuming a large uncertainty about the assumed parameters when constructing Bayesian designs for the multinomial logit model does lead to reasonably efficient designs for a mixed logit model. It is, however, still substantially better to take into account the respondent heterogeneity explicitly when constructing a design.

## Acknowledgments

The authors are grateful to the referees for their helpful suggestions and to Qi Ding for his contributions to the development of the algorithm. He was funded by project NB/06/05 of the National Bank of Belgium. In addition, the authors acknowledge the financial support of the research council of the Katholieke Universiteit Leuven (OT-project 04/07).

## Appendix

For generating fully Bayesian mixed logit designs, we assumed that the prior distribution that takes into account the uncertainty about the heterogeneity parameters $\sigma_i^2$ is an inverse gamma distribution with mean and variance both equal to one. Fifty Halton draws were taken from this prior distribution. The same systematic 20 *prior draws* used for constructing semi-Bayesian mixed logit designs were used for the prior of the mean parameter $\boldsymbol{\mu}_{\boldsymbol{\beta}}$. With these draws, we ran the coordinate-exchange algorithm 1,500 times, yielding a set of 1,500 candidate designs. We then used one million Monte Carlo draws (1,000 for $\boldsymbol{\mu}_{\boldsymbol{\beta}} \times 1{,}000$ for $\boldsymbol{\sigma}_{\boldsymbol{\beta}}$) to reevaluate the best 20 of these designs according to the $D_{\mathrm{BM}}$-errors obtained by the $20 \times 50$ sample. Restricting ourselves to reevaluating only 20 candidate designs was necessary because the computation of the $D_{\mathrm{BM}}$-error took approximately 11 hours for a single design.

## References

Allenby, G. M., N. Arora, J. L. Ginter. 1998. On the heterogeneity of demand. *J. Marketing Res.* **35**(3) 384–389.

Arora, N., T. Henderson. 2007. Embedded premium promotion: Why it works and how to make it more effective. *Marketing Sci.* **26**(4) 514–531.

Bhat, C. R. 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Res. Part B* **35**(7) 677–693.

Bunch, D. S., J. J. Louviere, D. Anderson. 1996. A comparison of experimental design strategies for multinomial logit models:The case of generic attributes. Working paper, Graduate School of Management, University of California, Davis, Davis.

Glasgow, G. 2001. Mixed logit models for multiparty elections. *Political Anal.* **9**(2) 116–136.

Haaijer, M. E., M. Vriens, T. Wansbeek, M. Wedel. 1998. Utility covariances and context effects in conjoint MNP models. *Marketing Sci.* **17**(3) 236–252.

Hensher, D. A., W. H. Greene. 2003. The mixed logit model: The state of practice. *Transportation* **30**(2) 133–176.

Huber, J., K. Train. 2001. On the similarity of classical and Bayesian estimates of individual mean partworths. *Marketing Lett.* **12**(3) 259–269.

Huber, J., K. Zwerina. 1996. The importance of utility balance in efficient choice design. *J. Marketing Res.* **33**(3) 307–317.

Kessels, R., P. Goos, M. Vandebroek. 2006. A comparison of criteria to design efficient choice experiments. *J. Marketing Res.* **43**(3) 409–419.

Kessels, R., B. Jones, P. Goos, M. Vandebroek. 2008. An efficient algorithm for constructing Bayesian optimal choice designs. *J. Bus. Econom. Statist.* Forthcoming.

Luo, L., P. K. Kannan, B. T. Ratchford. 2007. New product development under channel acceptance. *Marketing Sci.* **26**(2) 149–163.

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

McFadden, D., K. Train. 2000. Mixed MNL models of discrete response. *J. Appl. Econom.* **15**(5) 447–470.

Meyer, R. K., C. J. Nachtsheim. 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* **37**(1) 60–69.

Revelt, D., L. Train. 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Rev. Econom. Statist.* **80**(4) 647–657.

Sándor, Z., M. Wedel. 2001. Designing conjoint choice experiments using managers' prior beliefs. *J. Marketing Res.* **38**(4) 430–444.

Sándor, Z., M. Wedel. 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Sci.* **21**(4) 455–475.

Sándor, Z., M. Wedel. 2005. Heterogeneous conjoint choice designs. *J. Marketing Res.* **42**(2) 210–218.

Sloan, I. H., H. Wozniakowski. 1998. When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *J. Complexity* **14**(1) 1–33.

Toubia, O., J. R. Hauser. 2007. Research note—On managerially efficient experimental designs. *Marketing Sci.* **26**(6) 851–858.

Train, K. 2000. Halton sequences for mixed logit. Economics Working Papers E00-278, University of California, Berkeley, Berkeley.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

Wedel, M., W. A. Kamakura, N. Arora, A. Bemmaor, J. Chiangm, T. Elrod, R. Johnson, P. Lenk, S. Neslin, C. S. Poulsen. 1999. Discrete and continuous representation of heterogeneity. *Marketing Lett.* **10**(3) 217–230.

Woods, D. C., S. M. Lewis, J. A. Eccleston, K. G. Russell. 2006. Designs for generalized linear models with several variables and model uncertainty. *Technometrics* **48**(2) 284–292.

Zwerina, K., J. Huber, W. F. Kuhfeld. 1996. A general method for constructing efficient choice designs. Working paper, Fuqua School of Business, Duke University, Durham, NC.