

Efficient conjoint choice designs in the presence of respondent heterogeneity

Jie Yu, Peter Goos and Martina Vandebroek

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

Efficient Conjoint Choice Designs in the Presence of Respondent Heterogeneity

Jie Yu

Faculty of Economics and Applied Economics, Katholieke Universiteit Leuven,
Naamsestraat 69, B-3000 Leuven, Belgium.

Email: Jie.Yu@econ.kuleuven.be

Tel: +32 (0)16 32 69 62

Fax: +32 (0)16 32 67 32

Peter Goos

Faculty of Applied Economics, Universiteit Antwerpen, Prinsstraat 13, B-2000
Antwerpen, Belgium.

Email: Peter.Goos@ua.ac.be

Tel: +32 (0)3 220 40 59

Fax: +32 (0)3 220 48 17

Martina Vandebroek

Faculty of Economics and Applied Economics & University Center for Statistics,
Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.

Email: Martina.Vandebroek@econ.kuleuven.be

Tel: +32 (0)16 32 69 75

Fax: +32 (0)16 32 67 32

Efficient Conjoint Choice Designs in the Presence of Respondent Heterogeneity

Abstract

The authors propose a fast and efficient algorithm for constructing D -optimal conjoint choice designs for mixed logit models in the presence of respondent heterogeneity. With this new algorithm, the construction of semi-Bayesian D -optimal mixed logit designs with large numbers of attributes and attribute levels becomes practically feasible. The results from the comparison of eight designs (ranging from the simple locally D -optimal design for the multinomial logit model and the nearly orthogonal design generated by Sawtooth (CBC) to the complex semi-Bayesian mixed logit design) across wide ranges of parameter values show that the semi-Bayesian mixed logit approach outperforms the competing designs not only in terms of estimation efficiency but also in terms of prediction accuracy. In particular, it was found that semi-Bayesian mixed logit designs constructed with large heterogeneity parameters are most robust against the misspecification of the values for the mean of the individual-level coefficients for making precise estimations and predictions.

Keywords: semi-Bayesian mixed logit design, heterogeneity, prediction accuracy, multinomial logit design, model-robust design, D -optimality, algorithm

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; 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) 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 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 better quality data than the multinomial logit model because it offers an extended framework within which a greater amount of true behavioural 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 heterogeneous preferences across respondents (Huber and Zwerina 1996; Kessels et al. 2006a; 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 of the individual-level coefficients, denoted by

$\boldsymbol{\mu}_\beta$ in this article, and the covariances, $\boldsymbol{\Sigma}_\beta$, of these 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 adopted in this article.

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 propose a fast and efficient algorithm that reduces the computation time dramatically for generating efficient mixed logit designs. We extend the work of Sándor and Wedel (2002) by using a Bayesian approach to construct semi-Bayesian D -optimal designs for the mixed logit model which take into account the uncertainty about the mean parameters $\boldsymbol{\mu}_\beta$ in the design construction process. We 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. We also study the sensitivity of the semi-Bayesian mixed logit designs to the misspecification of $\boldsymbol{\Sigma}_\beta$, the heterogeneity parameters. In addition, we examine how 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 Section 3, we present the methodology to construct efficient designs for the mixed logit model in a computationally efficient way. In Section 4, we describe the details of the simulation study and the performance evaluation criteria. In Section 5, we evaluate the proposed designs in terms of the efficiency of parameter estimation and of the predictive performance. Section 6 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

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

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

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

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 $\boldsymbol{\mu}_\beta$ and diagonal covariance matrix $\boldsymbol{\Sigma}_\beta = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$. We call $\boldsymbol{\sigma}_\beta = (\sigma_1, \sigma_2, \dots, \sigma_p)'$ the heterogeneity vector as it captures the heterogeneity across respondents. The larger the values in $\boldsymbol{\sigma}_\beta$, the larger the degree of heterogeneity among respondents. Now, $\boldsymbol{\beta}$ can be written as $\boldsymbol{\beta} = \boldsymbol{\mu}_\beta + \mathbf{V}\boldsymbol{\sigma}_\beta$, where \mathbf{V} is a diagonal matrix having the random vector $\mathbf{v} = (v_1, v_2, \dots, v_p)'$ 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

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

where h represents the standard normal density function. In this article, we are interested in estimating the mean parameter vector $\boldsymbol{\mu}_\beta$ and the heterogeneity vector $\boldsymbol{\sigma}_\beta$ 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 $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\sigma}_\beta$.

2.2 Design Efficiency Criterion

A well-known criterion for evaluating the efficiency of experimental designs is the D -optimality criterion. 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 $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\sigma}_\beta$. The information matrix on $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\sigma}_\beta$, which is inversely proportional to the covariance matrix of the parameter estimates, is given by the $2p \times 2p$ dimensional matrix

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

where

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

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

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

$$D_{M\text{-error}} = \det\{\mathcal{I}(\boldsymbol{\mu}_\beta, \boldsymbol{\sigma}_\beta)^{-1}\}^{1/2p}. \quad (7)$$

In this paper, we adopt a Bayesian approach for constructing D -optimal designs for the mixed logit model. For reasons of computational convenience and ease of interpretation of the results, we only take into consideration the uncertainty about the mean parameter

vector $\boldsymbol{\mu}_\beta$, and not that about the heterogeneity vector $\boldsymbol{\sigma}_\beta$. We therefore use a fixed value for $\boldsymbol{\sigma}_\beta$ when we construct the designs and call the D -optimal designs derived in this paper semi-Bayesian rather than Bayesian D -optimal designs. In the simulation study discussed below, we investigate the robustness of the proposed design to the misspecification of the heterogeneity vector.

We denote the prior distribution for $\boldsymbol{\mu}_\beta$ by $g(\boldsymbol{\mu}_\beta)$. This distribution can be informative or uninformative depending on the amount of prior information available to the researcher. The semi-Bayesian D -optimal mixed logit designs are constructed by minimizing the D_{BM} -error, which is the expectation of the D_M -error over the prior distribution $g(\boldsymbol{\mu}_\beta)$:

$$D_{BM\text{-error}} = \int \det\{\mathcal{I}(\boldsymbol{\mu}_\beta, \boldsymbol{\sigma}_\beta)^{-1}\}^{1/2p} g(\boldsymbol{\mu}_\beta) d\boldsymbol{\mu}_\beta. \quad (8)$$

3 Design Construction Algorithm

As mentioned by Sándor and Wedel (2002), the large computation time for generating mixed logit designs makes it infeasible to construct Bayesian designs for realistic problems using their RSC-algorithm (Relabeling, Sampling and Cycling). In this section, we introduce a new design algorithm, which we call the Adaptive Halton algorithm (AH), for generating semi-Bayesian D -optimal mixed logit designs. This algorithm overcomes the limitation of previous algorithms for constructing mixed logit designs as it 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 utilized in the construction of a semi-Bayesian D -optimal mixed logit design. Each of them are 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, the large numbers of random draws needed for approximating these integrals lead to long computation times. In our new algorithm, we therefore make systematic draws from the distributions $f(\boldsymbol{\beta})$ and $g(\boldsymbol{\mu}_\beta)$ 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 AH algorithm has four features which speed up the computation of semi-Bayesian D -optimal mixed logit designs: (i) the coordinate-exchange procedure, (ii) the fast update of the information matrix, (iii) the small sample of *mixed logit draws*, and (iv) the small sample of *prior draws*. Each of these improvements is described in detail in the following sections.

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. (2006b). Unlike the algorithm in Kessels et al. (2006a), the coordinate-exchange algorithm does not require the construction of a candidate set. 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. 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_{BM} -error. The first iteration is terminated once 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 1000 different starting designs to find the designs reported in this paper.

3.2 Updating the Information Matrix

A traditional way to update the information matrix is to recompute the information matrix for all choice sets once an attribute level in one set is changed. This is not necessary when updating the information matrix in formula (4) since it is the sum of the per choice set information matrices. In fact, changing an attribute level in a profile of the s th choice set only affects the choice set \mathbf{X}_s , but not the rest of the choice sets. This provides a computational shortcut to update the information matrix. For simplicity, we define the new way of updating the information matrix as a single-set update and the traditional way as a whole-set update. The updated information matrix \mathcal{I}_{update} for the new updating approach can be written as follows:

$$\mathcal{I}_{update} = \mathcal{I}_{old} - \mathcal{I}_s^{old} + \mathcal{I}_s^{new} \quad (9)$$

where \mathcal{I}_{old} is the information matrix for the whole design before the profile in set s is changed, and \mathcal{I}_s^{old} and \mathcal{I}_s^{new} are the information matrices for set s before and after the change, respectively. In this way, for each exchange, we only need an update of the information matrix for the choice set to which the modified profile belongs. This leads to a much faster algorithm. For example, for a conjoint choice design with 12 choice sets, using the single-set update instead of a whole-set update leads to a saving of 91.7% of the computation time.

3.3 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). This is confirmed by Train (2000) who showed that the estimation of a mixed logit model can require 2-3 hours for moderately sized models, and that run times of 10-20 hours are common.

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, Halton sequences lead to a faster convergence and smaller simulation errors (Bhat 2001). As a precise approximation of the information matrix for a mixed logit model is crucial for the purpose of generating

semi-Bayesian D -optimal designs, we used as many as 250 Halton draws. This led to better results than 1000 random Monte Carlo draws while saving 75% of the computing time. The detailed construction procedures and the reason why Halton draws perform better than random draws can be found in Train (2000).

3.4 Small Numbers of Prior Draws

The integral that has to be computed for evaluating the D_{BM} -error in (8) does not require as many as 250 draws. Our own computational work as well as that described in Kessels et al. (2006b) suggests that a designed sample of 20 *prior draws* instead 1000 Monte Carlo draws yields satisfactory results. The designed sample of 20 *prior draws* only provides 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 implementation of a large number of runs of the coordinate-exchange algorithm for searching the best design within a reasonable time. Kessels et al. (2006b) show that with the small sample of intelligent draws and more runs in the design construction, they are able to find more efficient designs than by using the large random Monte Carlo sample. The methodology they used 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 procedures for constructing the 20-point sample and for checking its performance can be found in Kessels et al. (2006b). The key is to generate 20 prior parameter vectors that are uniformly distributed on a hypersphere. To evaluate the effectiveness of the systematic 20-point set of draws, we have compared D_{BM} -errors computed using the random Monte Carlo sample to D_{BM} -errors computed using the systematic sample for several designs. The results are visualized in Figures 1 and 2.

The plot in Figure 1 shows how an initial random design is improved by making a number of changes to it. Each asterisk corresponds to a design that is found with improved D_{BM} -error during the coordinate-exchange algorithm. The initial random design is represented by the asterisk in the upper right hand corner of the picture, while the best design ultimately produced by the algorithm is represented by the asterisk in the lower left hand corner. The pattern in the asterisks shows that, for an improvement in the D_{BM} -error computed using the systematic 20-point sample, there is a matching decrease in the D_{BM} -error computed more accurately using the 1000-point Monte Carlo sample. This observation is supported by the fact that the D_{BM} -error values produced by the two methods have a correlation of 99.84%.

Figure 2 compares the two methods for approximating the D_{BM} -error in another way. Each asterisk in Figure 2 represents the best design found with a single random start of the coordinate-exchange algorithm. The correlation between the D_{BM} -error values for the systematic and the Monte Carlo sample is now 94.3%. This implies that there is a high chance that an efficient design constructed using the systematic 20-point sample is also efficient when it is evaluated by the random Monte Carlo sample.

Due to the facts that the systematic 20-point approximation method reduces the computation time considerably and that the two approximation methods almost perfectly

agree with each other in searching efficient mixed logit designs, it is justified to use the systematic 20-point set of *prior draws*. However, we follow the suggestion in Kessels et al. (2006b) to use the small systematic sample in the coordinate-exchange procedure and to re-evaluate the designs that procedure yields by means of the larger Monte Carlo sample. This is why our algorithm is named an adaptive algorithm.

3.5 Comparing the AH algorithm and a Benchmark Algorithm in Terms of Computation Time

To demonstrate how fast the AH algorithm is, we compared it to a benchmark algorithm in terms of computation time. The benchmark algorithm used random samples of 1000 *prior draws* and 1000 *mixed logit draws*, and it utilized the traditional whole-set update procedure for the information matrix. The features of the benchmark algorithm and the AH algorithm are summarized in Table 1. We recorded the computation times for one try of both algorithms for three design settings: $3^4/2/18$, $3^4/3/12$ and $3^4/4/9$. Each of these settings involves four attributes with three levels. The first setting has 18 choice sets of two profiles, while the second and third settings correspond to choice designs with 12 choice sets of three profiles and nine choice sets of four profiles, respectively. The computations were made by a Dell PC with two 2.8 GHz Intel Processors and 1GB RAM. The computing times, expressed in hours:minutes:seconds, are shown in Table 2.

Table 1: Features of the benchmark algorithm and the AH algorithm

	Benchmark	Adaptive Halton
Exchange algorithm	Coordinate-exchange	Coordinate-exchange
Number of <i>prior draws</i>	1000	20
Number of <i>mixed logit draws</i>	1000	250
Information matrix	whole-set update	single-set update

Table 2: Computing times for the benchmark algorithm and the AH algorithm

	$3^4/2/18$	$3^4/3/12$	$3^4/4/9$
AH algorithm	00:03:26	00:04:16	00:05:43
Benchmark algorithm	61:39:24	65:02:46	70:28:10

The table shows that the computing times for the AH algorithm are much smaller than those for the benchmark algorithm. The computing times for the AH algorithm are such that the construction of semi-Bayesian D -optimal designs for the mixed logit model becomes practically feasible. In particular, with the newly developed algorithm, the computation of semi-Bayesian D -optimal designs for the mixed logit model is no longer restricted to small design problems.

4 Design Evaluation

In this section, we compare the performance of eight types of designs under different conditions. We investigate which designs allow for an efficient parameter estimation and accurate predictions when there is heterogeneity in individual-level coefficients across respondents and when information about the mean parameter values 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 3. In the table, $\mathbf{1}_8$ and $\mathbf{0}_8$ denote an 8×1 vector of ones and an 8×1 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 $g(\boldsymbol{\mu}_\beta)$ for $\boldsymbol{\mu}_\beta$ but with a different prior value for the heterogeneity vector $\boldsymbol{\sigma}_\beta$. The prior distribution $g(\boldsymbol{\mu}_\beta)$ 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 8-dimensional identity matrix. The elements of the heterogeneity vector $\boldsymbol{\sigma}_\beta$ were 0.5, 1 and 1.5 for the three designs. 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, $\boldsymbol{\sigma}_\beta = \mathbf{1}_8$, 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.

Table 3: 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	-	-

Furthermore, two Bayesian D -optimal designs for a multinomial logit model, which assumes respondent homogeneity, were used in the study. These two designs were computed using the prior mean vector $\boldsymbol{\mu}_0$ and variance-covariance matrices \mathbf{I}_8 and $9\mathbf{I}_8$. 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.

4.2 Details of the Simulation Study

In our simulation study, we evaluated the eight designs in Table 3 under five conditions. In each case, a different parameter space Ω_i was used to draw 1000 values of the mean vector μ_β . The heterogeneity vector σ_β was fixed at $\mathbf{1}_8$ for each condition. In the sequel of this article, we denote the r th draw for a given parameter space by μ_β^r .

For the parameter spaces Ω_1 , Ω_2 and Ω_3 , the mean parameter vector μ_β , used for the data generation, was drawn from distributions centered around μ_0 with covariance matrices $0.25\mathbf{I}_8$, \mathbf{I}_8 and $2.25\mathbf{I}_8$, respectively. As a consequence, the parameter spaces Ω_1 , Ω_2 and Ω_3 are characterized by an increasing degree of mean parameter misspecification. For the parameter spaces Ω_4 and Ω_5 , the mean parameter vector μ_β was drawn from distributions with a mean that is different from μ_0 and with covariance matrix \mathbf{I}_8 . For parameter space Ω_4 , the mean vector was $-\mu_0$, while for parameter space Ω_5 it was $-\mu_0 + 0.5\mathbf{1}_8$. The degree of mean vector misspecification is therefore even larger for these parameter spaces than for Ω_1 , Ω_2 and Ω_3 . An overview of the five parameter spaces is given in Table 4.

Table 4: Parameter space specifications

Ω_1	$\mu_\beta \sim N(\mu_0, 0.25\mathbf{I}_8)$
Ω_2	$\mu_\beta \sim N(\mu_0, \mathbf{I}_8)$
Ω_3	$\mu_\beta \sim N(\mu_0, 2.25\mathbf{I}_8)$
Ω_4	$\mu_\beta \sim N(-\mu_0, \mathbf{I}_8)$
Ω_5	$\mu_\beta \sim N(-\mu_0 + 0.5\mathbf{1}_8, \mathbf{I}_8)$

4.3 Performance Criteria

In our simulation study, we have used four different criteria to evaluate the eight different design options. The first two of these quantify the precision of the model estimation. The other two criteria measure the accuracy of the predictions made.

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 true mean parameter μ_β^r , while fixing the heterogeneity vector σ_β at $\mathbf{1}_8$. We denote that locally optimal design by $\mathbf{X}_{\mu_\beta^r}$. For a given μ_β^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 $\mathbf{X}_{\mu_\beta^r}$ assuming that $\mu_\beta = \mu_\beta^r$:

$$RLD(\mathbf{X}, \mu_\beta^r) = \frac{D_{M\text{-error}}(\mathbf{X}_{\mu_\beta^r} | \mu_\beta = \mu_\beta^r)}{D_{M\text{-error}}(\mathbf{X} | \mu_\beta = \mu_\beta^r)}. \quad (10)$$

This expression lies between zero and one. $RLD(\mathbf{X}, \mu_\beta^r)$ -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 $\mu_\beta = \mu_\beta^r$. Designs that have $RLD(\mathbf{X}, \mu_\beta^r)$ -efficiencies close to one for all draws μ_β^r from a certain parameter space Ω_i are desirable.

In our simulation study, we computed $RLD(\mathbf{X}, \boldsymbol{\mu}_\beta^r)$ -efficiencies for every draw $\boldsymbol{\mu}_\beta^r$ for each of the five parameter spaces. Histograms that visualize the distribution of the resulting 1000 $RLD(\mathbf{X}, \boldsymbol{\mu}_\beta^r)$ -efficiencies for each of the eight design options in Table 3 are shown in Figures 3-7. These results are discussed in detail in Section 5.1.

4.3.2 Percentage Reduction in the Number of Observations

To assess the relative performance of a pair of designs, we also computed the percentage 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 is defined by

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

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

4.3.3 Expected Root Mean-Squared Prediction Error

To assess the predictive accuracy of the constructed designs, we computed the expected root mean-squared prediction error ($ERMSE_P$) for the five parameter spaces defined in Table 4. The $ERMSE_P$ compares the predicted to the true probabilities based on a $3^4/2/12$ holdout design generated using Sawtooth and is defined as

$$ERMSE_P = \int [(\pi(\hat{\boldsymbol{\mu}}_\beta, \hat{\boldsymbol{\sigma}}_\beta) - \pi(\boldsymbol{\mu}_\beta, \boldsymbol{\sigma}_\beta))' (\pi(\hat{\boldsymbol{\mu}}_\beta, \hat{\boldsymbol{\sigma}}_\beta) - \pi(\boldsymbol{\mu}_\beta, \boldsymbol{\sigma}_\beta))]^{1/2} \phi(\hat{\boldsymbol{\mu}}_\beta, \hat{\boldsymbol{\sigma}}_\beta) d\hat{\boldsymbol{\mu}}_\beta d\hat{\boldsymbol{\sigma}}_\beta, \quad (12)$$

where $\pi(\hat{\boldsymbol{\mu}}_\beta, \hat{\boldsymbol{\sigma}}_\beta)$ is the vector containing the predicted probabilities computed using the parameter estimates $\hat{\boldsymbol{\mu}}_\beta$ and $\hat{\boldsymbol{\sigma}}_\beta$, $\pi(\boldsymbol{\mu}_\beta, \boldsymbol{\sigma}_\beta)$ is the vector of the probabilities obtained using the true parameter vectors $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\sigma}_\beta$, and $\phi(\hat{\boldsymbol{\mu}}_\beta, \hat{\boldsymbol{\sigma}}_\beta)$ is the asymptotic distribution of the parameter estimates. For a design \mathbf{X} and true parameter vectors $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\sigma}_\beta$, the asymptotic distribution $\phi(\hat{\boldsymbol{\mu}}_\beta, \hat{\boldsymbol{\sigma}}_\beta)$ of the parameter estimates is the multivariate normal distribution with mean vector $(\boldsymbol{\mu}_\beta', \boldsymbol{\sigma}_\beta)'$ and covariance matrix $\mathcal{I}(\boldsymbol{\mu}_\beta, \boldsymbol{\sigma}_\beta | \mathbf{X})^{-1}$. We approximated the integral in (12) by using 1000 random draws from that asymptotic distribution, and we computed $ERMSE_P$ values for 1000 draws $\boldsymbol{\mu}_\beta^r$ from each of the five parameter spaces for the eight designs in Table 3. Histograms of the $ERMSE_P$ values for parameter space Ω_2 are shown in Figure 8 and discussed in Section 5.2, along with the results for the other parameter spaces.

4.3.4 Percentage Decrease in Prediction Error

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

$$1 - \frac{ERMSE_P(\mathbf{X}_1 | \boldsymbol{\mu}_\beta = \boldsymbol{\mu}_\beta^r)}{ERMSE_P(\mathbf{X}_2 | \boldsymbol{\mu}_\beta = \boldsymbol{\mu}_\beta^r)} \quad (13)$$

for a given draw $\boldsymbol{\mu}_\beta^r$. 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 3 in terms of the efficiency of parameter estimation when the true mean parameters $\boldsymbol{\mu}_\beta$ are taken from the five parameter spaces in Table 4. The aim is to find out which design option, in general, is more robust against the misspecification of the mean parameter values. Figures 3-7 present the distributions of the *RLD*-efficiencies of the eight designs for each of the five parameter spaces. Values close to one indicate that the design is almost as good as the locally optimal design. Four common features occur across these five figures.

First of all, compared to the designs that were constructed ignoring the respondent heterogeneity (see the plots (e)-(h) in each of the Figures 3-7), the optimal designs for the mixed logit model (see the plots (a)-(d) in each figure) perform substantially better. This holds in particular for parameter space Ω_1 , where the mean parameter vectors $\boldsymbol{\mu}_\beta^r$ are all close to the one assumed when constructing the optimal designs, $\boldsymbol{\mu}_0$. This can be seen in Figure 3. Furthermore, the semi-Bayesian *D*-optimal mixed logit designs constructed with heterogeneity parameter vector $\boldsymbol{\sigma}_\beta = \mathbf{1}_8$ and with overspecified heterogeneity vector $\boldsymbol{\sigma}_\beta = 1.5\mathbf{1}_8$ tend to 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.

Secondly, comparing the semi-Bayesian *D*-optimal designs for the mixed logit model constructed with $\boldsymbol{\sigma}_\beta = 1.5\mathbf{1}_8$ and with $\boldsymbol{\sigma}_\beta = 0.5\mathbf{1}_8$ to the one obtained using $\boldsymbol{\sigma}_\beta = \mathbf{1}_8$ shows that overspecifying the heterogeneity vector $\boldsymbol{\sigma}_\beta$ does not have a large negative impact on the efficiency of parameter estimation. For the parameter spaces Ω_1 , Ω_2 , Ω_3 and Ω_4 , the semi-Bayesian mixed logit design constructed using $\boldsymbol{\sigma}_\beta = \mathbf{1}_8$ leads to reductions in the number of observations of only 2.01%, 3.4%, 2.4% and 0.49%, respectively, compared to the design obtained using the overspecified $\boldsymbol{\sigma}_\beta$. For parameter space Ω_5 , where the degree of misspecification of the mean parameter vector is considerably larger than for the other parameter spaces, the reduction is negative. The negative value for the reduction, -3.8%, demonstrates that the semi-Bayesian mixed logit design constructed by assuming a large heterogeneity vector is more robust to the misspecification of the mean parameters than when assuming a small heterogeneity. This result is very useful when prior information about the mean parameter vector $\boldsymbol{\mu}_\beta$ 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. In addition, we observe that the loss in efficiency for underspecifying $\boldsymbol{\sigma}_\beta$ is higher than that for overspecifying $\boldsymbol{\sigma}_\beta$. However, the design constructed with underspecified $\boldsymbol{\sigma}_\beta$ is still more efficient than the designs constructed for the multinomial logit model and which thereby ignore the respondent heterogeneity.

Thirdly, comparing the *RLD*-efficiencies of the semi-Bayesian *D*-optimal mixed logit design constructed using $\boldsymbol{\sigma}_\beta = \mathbf{1}_8$ in panel (b) of the Figures 3-7 to those of the locally *D*-optimal mixed logit design obtained using the same heterogeneity vector in panel (d) demonstrates the benefit of using the semi-Bayesian approach. For each parameter space Ω_i , it turns out that the semi-Bayesian design outperforms the locally optimal design. On average, the semi-Bayesian mixed logit designs require 3.3%, 9.1% and 11.5% fewer

observations to have the same efficiency as the locally D -optimal mixed logit design for the parameter spaces Ω_1 , Ω_2 and Ω_3 , respectively. For the parameter spaces Ω_4 and Ω_5 , which are centered away from the mean parameter vector μ_0 used during the design construction, the average reductions are 10% and 9.8%, respectively.

Fourthly, the locally D -optimal design for the multinomial logit model and the nearly orthogonal design, the efficiencies of which are displayed in the panels (g) and (h) of the Figure 3-7, yield substantially poorer RLD -efficiencies than the other design options. This is particularly true when they are compared to the semi-Bayesian mixed logit designs. For instance, for parameter space Ω_1 , the best and the worst RLD -efficiencies for the semi-Bayesian D -optimal mixed logit design constructed with $\sigma_\beta = \mathbf{1}_8$ are 96.4% and 57.1%, respectively (see panel (b) of Figure 3). In contrast, the best and worst RLD -efficiencies for the nearly orthogonal design are only 36.9% and 17.3%, respectively (see panel (h)). For the locally D -optimal design for the multinomial model, these values are 38.1% and 17.4%, respectively (see panel (g)). The maximum RLD -efficiencies of these two designs are thus much smaller than the minimum RLD -efficiency of the best semi-Bayesian D -optimal design for the mixed logit model. In addition, for a parameter space that accommodates a wide variety of mean parameters such as Ω_3 , the minimum RLD -efficiency is 0.009 for the locally D -optimal multinomial logit design and 0.020 for nearly orthogonal design. This indicates that, for some parameter values, almost no information is available for parameter estimation when these two designs are used. Another striking result is that 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 66.6% and 65.4%, respectively. This demonstrates 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 heterogeneous numbers of attribute levels, such as designs of the type $2 \times 3 \times 4/2/24$, and designs of the type $2 \times 2 \times 3/3/12$. The results for these studies did not reveal different insights in the relative performance of the eight design options studied, which is why we do not discuss them here.

5.2 Predictive Accuracy

In this section, we evaluate the eight designs in Table 3 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 parameter space Ω_2 . The results for the other four parameter spaces are similar and they are only briefly discussed at the end of this section. Figure 8 shows the $ERMSE_P$ -values of the eight designs for parameter space Ω_2 . Note that small $ERMSE_P$ values are desirable.

As can be seen in Figure 8, the semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$ generally performs best in terms of predictive accuracy. Compared to the locally D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$, adopting a semi-Bayesian approach in the design construction, on average, leads to a 17.5% 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 σ_β on the predictive accuracy can be examined by comparing the semi-Bayesian D -optimal mixed logit designs for $\sigma_\beta = 1.5\mathbf{1}_8$ and $\sigma_\beta = 0.5\mathbf{1}_8$ to that for $\sigma_\beta = \mathbf{1}_8$. As can be seen from the panels (a)-(c) in Figure 8, the effects are similar to those for the RLD -efficiencies in the previous section. The design constructed using the correctly specified heterogeneity vector σ_β produces slightly smaller $ERMSE_P$ values than the design with overspecified σ_β . The average decrease in prediction error is only 3.15%. Underspecifying σ_β leads to a larger loss in predictive accuracy than overspecifying it. Therefore, we can conclude that it is best to use large values for σ_β to construct designs if there is uncertainty about these parameters.

By looking at the panels (g) and (h) of Figure 8, 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 $ERMSE_P$ values and lead to large $ERMSE_P$ values on average. Therefore, they 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 63.1% and a 62.4% decrease in prediction error.

A comparison between the $ERMSE_P$ 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, 14% 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 the loss in prediction accuracy caused by ignoring the respondent heterogeneity.

Evidently, we performed similar studies for the other parameter spaces. The results are similar to those for Ω_2 except that, for parameter spaces Ω_4 and Ω_5 , the semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 1.5\mathbf{1}_8$ has the best predictive performance. As indicated in Table 4, Ω_4 and Ω_5 are centered away from the mean parameter vector μ_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 spaces. The best predictive performance by the semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 1.5\mathbf{1}_8$ again 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 to specify large values for the heterogeneity vector in the design construction algorithm when sufficient information on the true mean parameter values is lacking.

6 Conclusions

In this paper, we proposed a fast and efficient algorithm for constructing semi-Bayesian D -optimal designs for the mixed logit model. A key feature of the algorithm is that each integral involved in the computation of the Bayesian D -optimality criterion is approxi-

mated using a small number of intelligent draws instead of a large number of random Monte Carlo draws. With the algorithm, we are able to construct large choice designs for the mixed logit model and to incorporate the uncertainty about the model parameters in the design construction. 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 prediction accuracy in the presence of respondent heterogeneity. The simulation studies lead to the conclusion that the semi-Bayesian D -optimal designs for the mixed logit model consistently perform better than the other designs. Compared to these designs, using nearly orthogonal designs and locally D -optimal designs for a multinomial logit model, and thereby ignoring both 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. It also turns out that specifying large values for the heterogeneity parameters is the most robust strategy against misspecification of the mean parameter values as this will usually provide the most efficient parameter estimators as well as the best predictive accuracy. 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.

Acknowledgment

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

References

- Allenby, G. M., A. Neeraj, J. L. Ginter. 1998. On the heterogeneity of demand. *Journal of Marketing Research* 35(August) 384-389.
- Bhat, C. R. 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research 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, CA.
- Glasgow, G. 2001. Mixed logit models for multiparty elections. *Political Analysis* 9(2) 116-136.
- Haaijer, M.E., M. Vriens, T. Wansbeek, M. Wedel. 1998. Utility covariances and context effects in conjoint MNP models. *Marketing Science* 17(3): 236-252.
- Hensher, D. A., W. H. Greene. 2003. The mixed logit model: The state of practice. *Transportation* 30 133-176.

- Huber, J., K. Train. 2001. On the similarity of classical and Bayesian estimates of individual mean partworths. *Marketing Letters* 12(3) 259-269.
- Huber, J., K. Zwerina. 1996. The importance of utility balance in efficient choice design. *Journal of Marketing Research* 33(August) 307-317.
- Kessels, R., P. Goos, M. Vandebroek. 2006a. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research* 43(3) 409-419.
- Kessels, R., B. Jones, P. Goos, M. Vandebroek. 2006b. An efficient algorithm for constructing Bayesian optimal choice designs. Research Report KBI_0616, Faculty of Economics and Applied Economics, Katholieke Universiteit Leuven.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, ed. Paul Zarembka, New York: Academic Press 105-142.
- McFadden, D., K. Train. 2000. Mixed MNL models of discrete response. *Journal of Applied Economics* 15(5) 447-470.
- Meyer, R. K., C. J. Nachtsheim. 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* 37 60-69.
- Revelt, D., L. Train. 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics* 80 (November) 647-657.
- Sándor, Z., M. Wedel. 2001. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research* 38 (November) 430-444.
- Sándor, Z., M. Wedel. 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Science* 21 (Fall) 455-457.
- Sloan, I. H., H. Wozniakowski. 1998. When are quasi-monte carlo algorithms efficient for high dimensional integrals? *Journal of Complexity* 14 1-33.
- Train, K. 2000. Halton sequences for mixed logit. Economics Working Papers, E00-278, University of California at Berkeley.
- Train, K. 2003. Discrete choice methods with simulation. Cambridge University Press.
- 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 Letters* 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* 48 284-292.
- 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 27708.

Appendix A

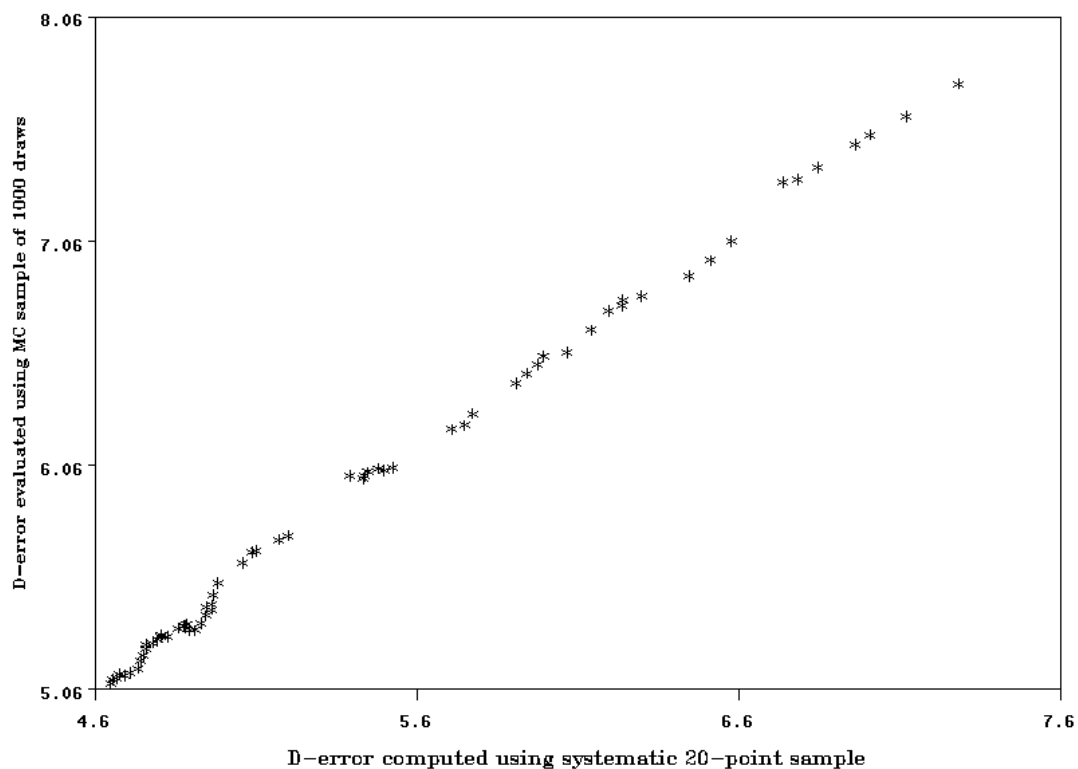


Figure 1: Evaluating the performance of the systematic 20-point sample by comparing its D_{BM} -error values to the corresponding D_{BM} -error values for the random Monte Carlo sample. The points describe the process of improvement within one random starting design by using the 20-point sample.

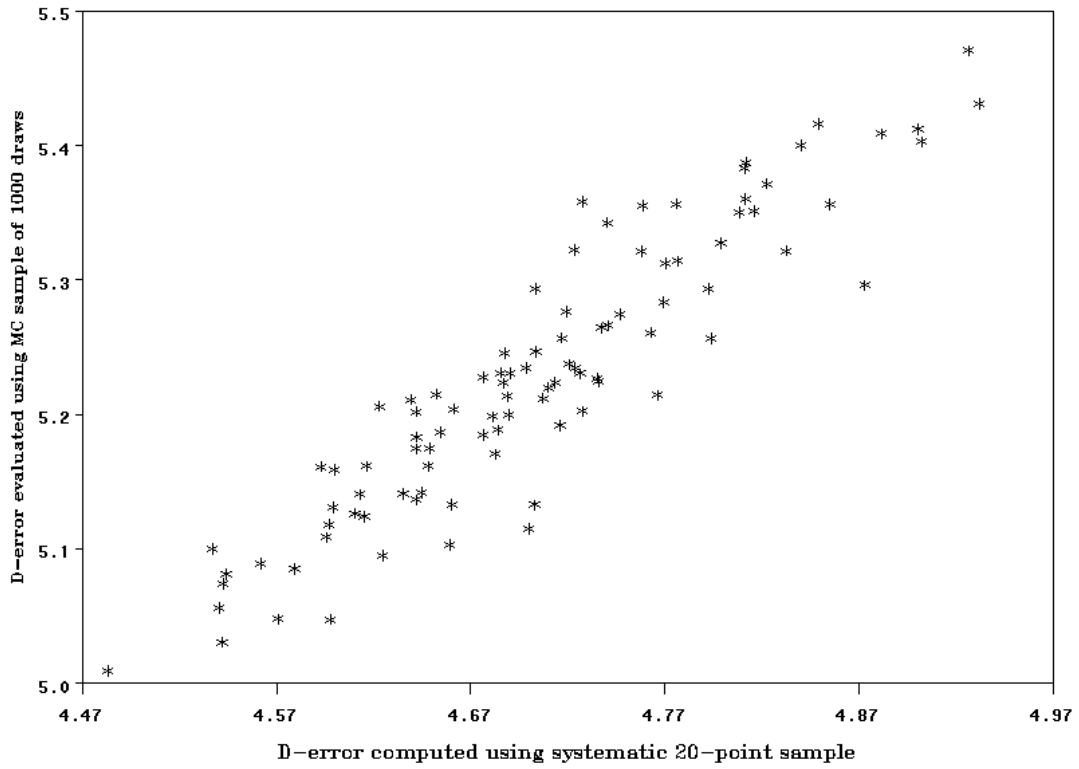
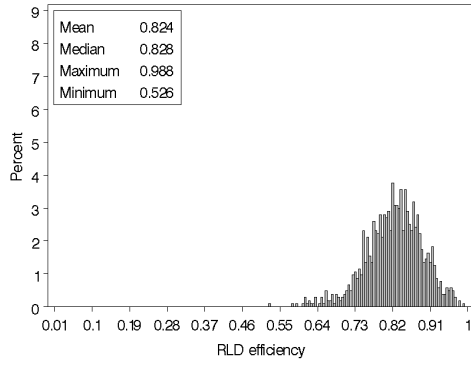
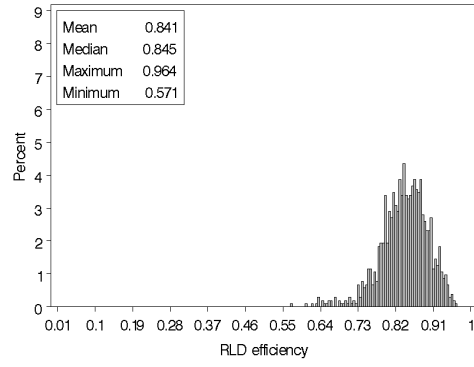


Figure 2: Evaluating the performance of the systematic 20-point sample by comparing its D_{BM} -error values to the corresponding D_{BM} -error values for the random Monte Carlo sample. Different points in this figure represent the best designs found with different random starting designs by using the 20-point sample.

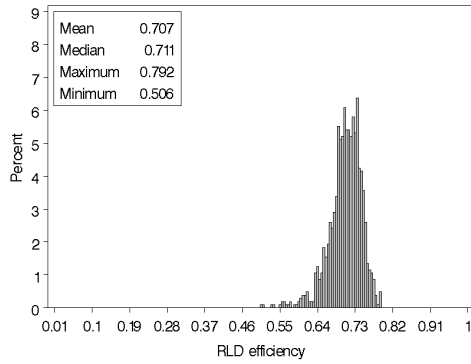
Appendix B



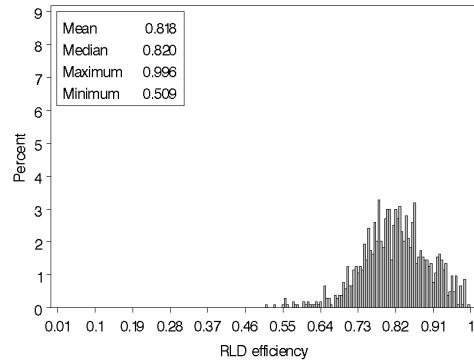
(a) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 1.51_8$



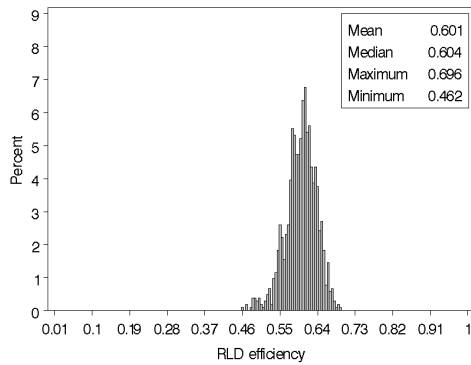
(b) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



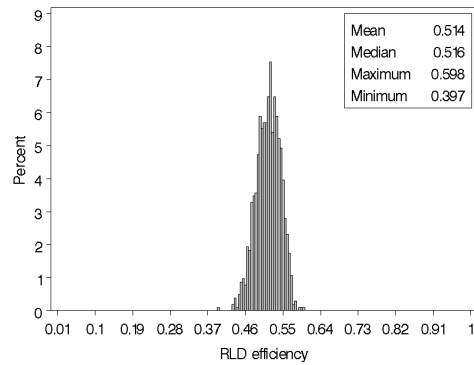
(c) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 0.51_8$



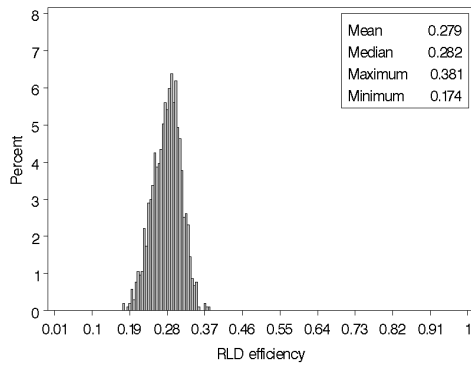
(d) Locally D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



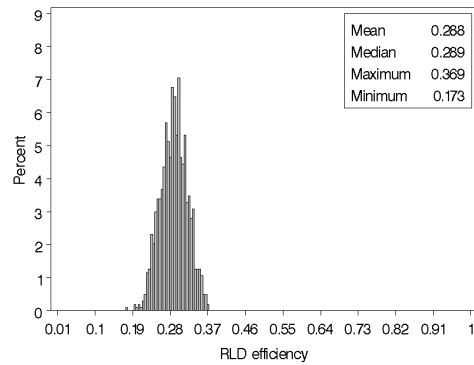
(e) Bayesian D -optimal multinomial logit design with covariance $9\mathbf{I}_8$



(f) Bayesian D -optimal multinomial logit design with covariance \mathbf{I}_8

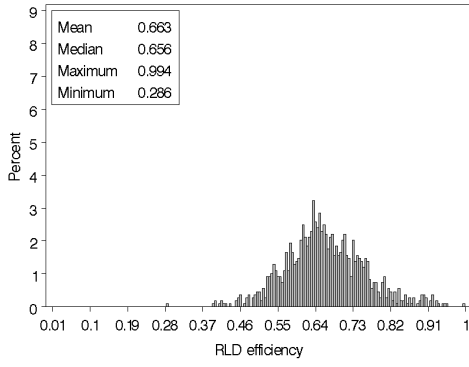


(g) Locally D -optimal multinomial logit design

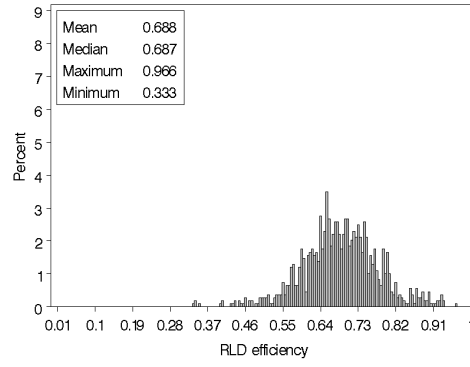


(h) Nearly orthogonal design

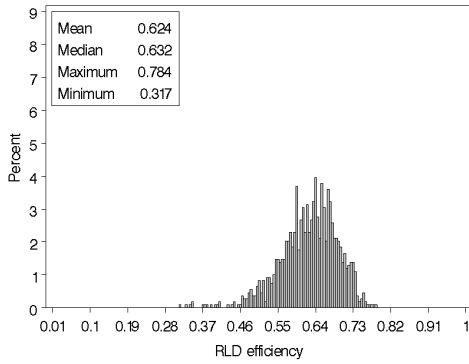
Figure 3: RLD -efficiencies for parameter space Ω_1 .



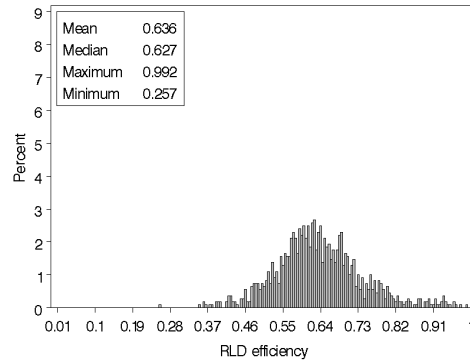
(a) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 1.51_8$



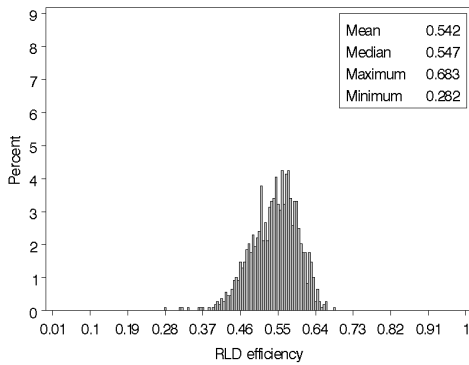
(b) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



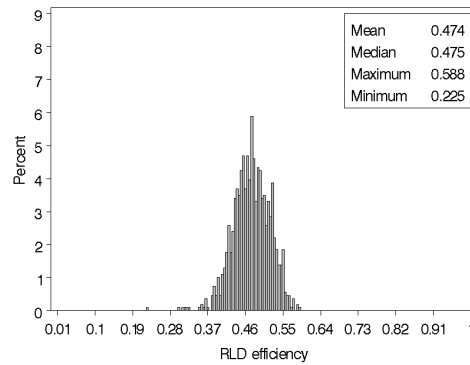
(c) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 0.51_8$



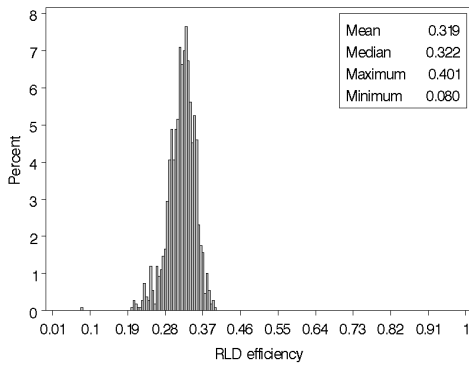
(d) Locally D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



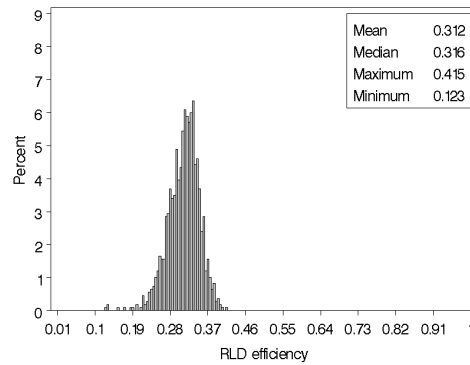
(e) Bayesian D -optimal multinomial logit design with covariance $9\mathbf{I}_8$



(f) Bayesian D -optimal multinomial logit design with covariance \mathbf{I}_8

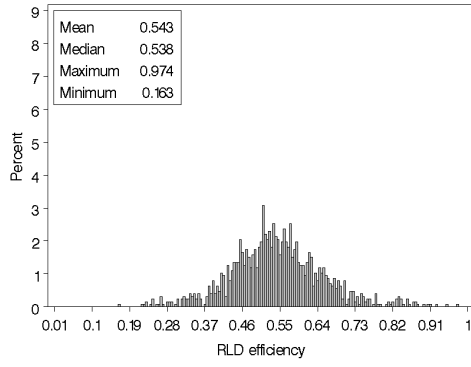


(g) Locally D -optimal multinomial logit design

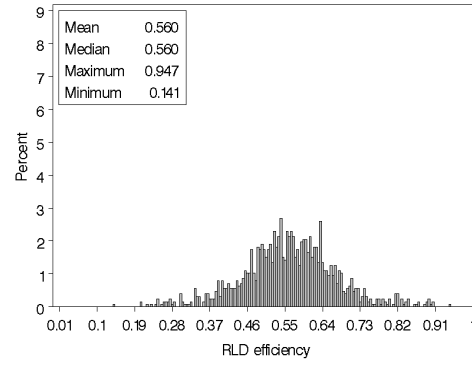


(h) Nearly orthogonal design

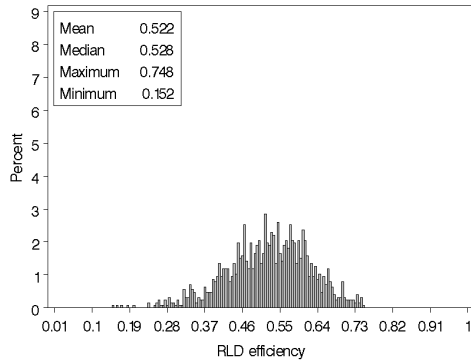
Figure 4: RLD -efficiencies for parameter space Ω_2 .



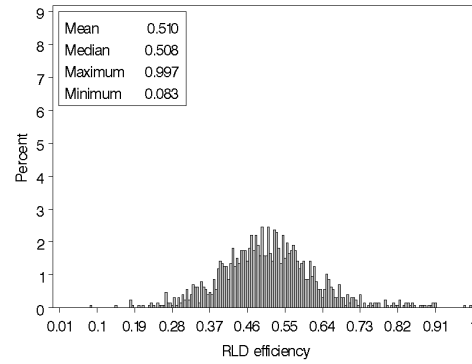
(a) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 1.51_8$



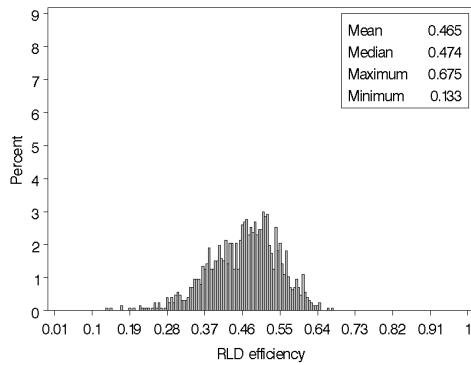
(b) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



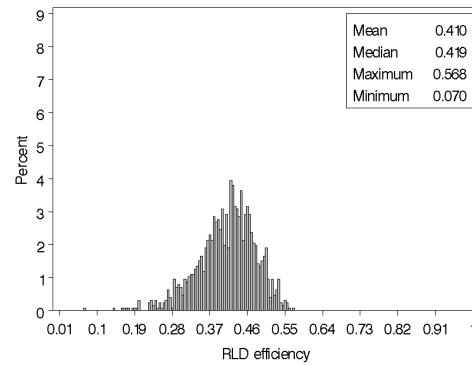
(c) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 0.51_8$



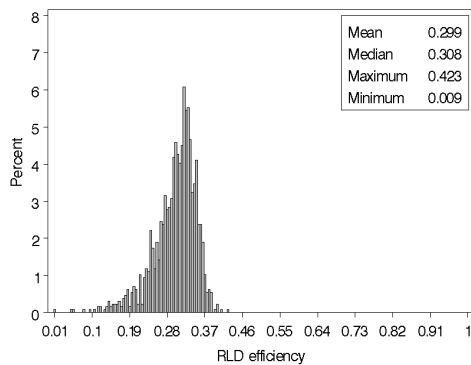
(d) Locally D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



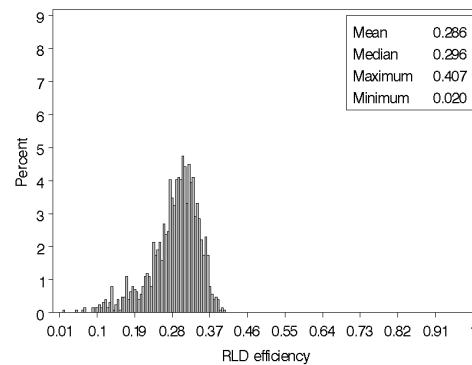
(e) Bayesian D -optimal multinomial logit design with covariance $9\mathbf{I}_8$



(f) Bayesian D -optimal multinomial logit design with covariance \mathbf{I}_8

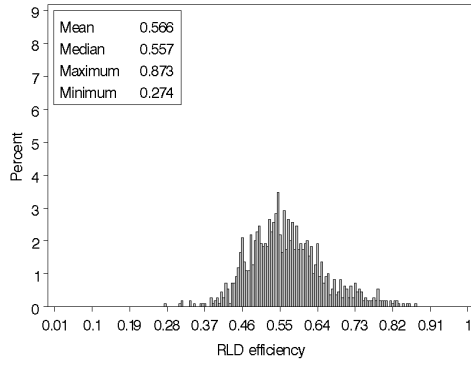


(g) Locally D -optimal multinomial logit design

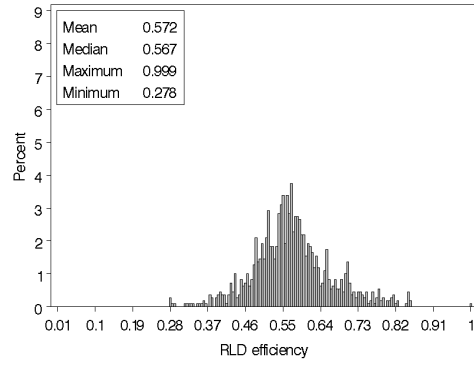


(h) Nearly orthogonal design

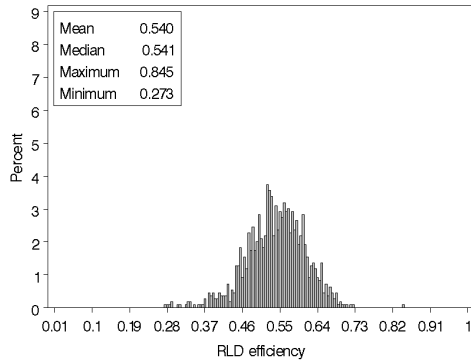
Figure 5: RLD -efficiencies for parameter space Ω_3 .



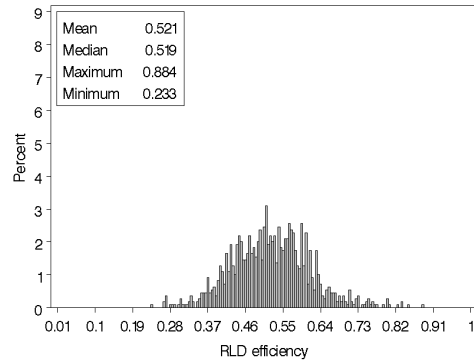
(a) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 1.51_8$



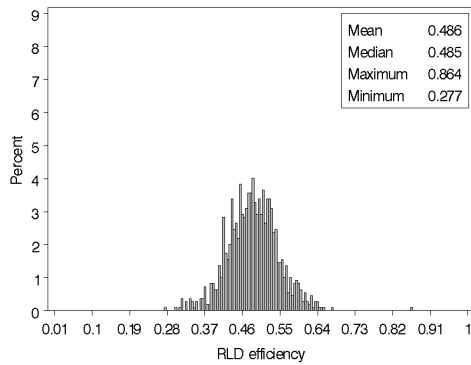
(b) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



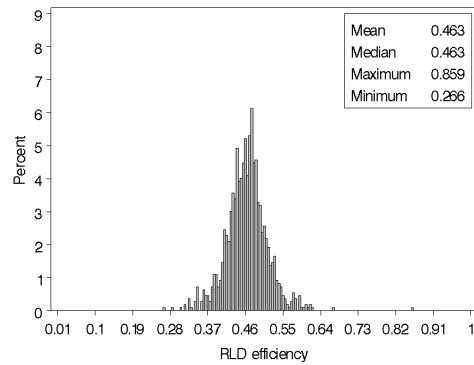
(c) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 0.51_8$



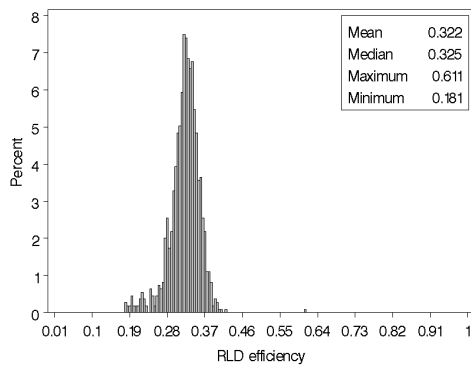
(d) Locally D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



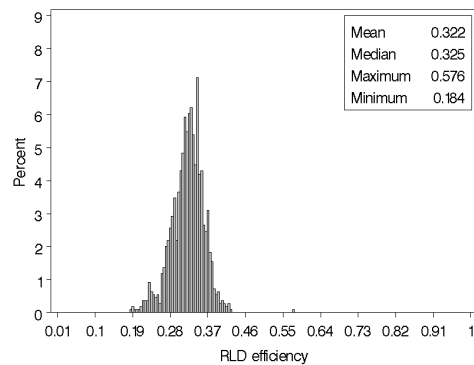
(e) Bayesian D -optimal multinomial logit design with covariance $9\mathbf{I}_8$



(f) Bayesian D -optimal multinomial logit design with covariance \mathbf{I}_8

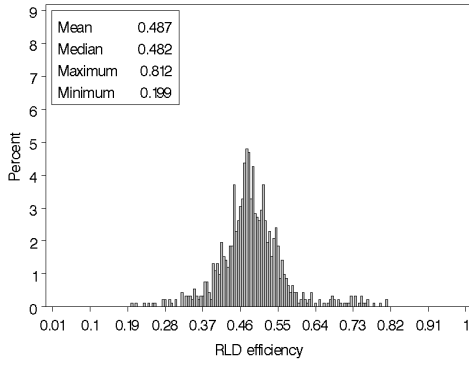


(g) Locally D -optimal multinomial logit design

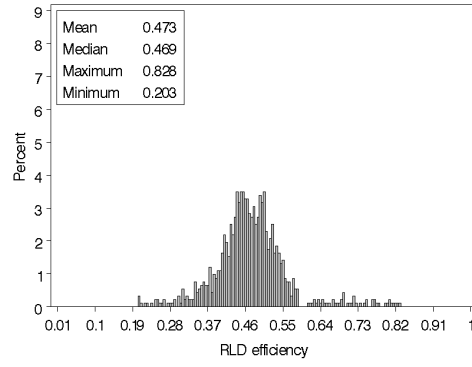


(h) Nearly orthogonal design

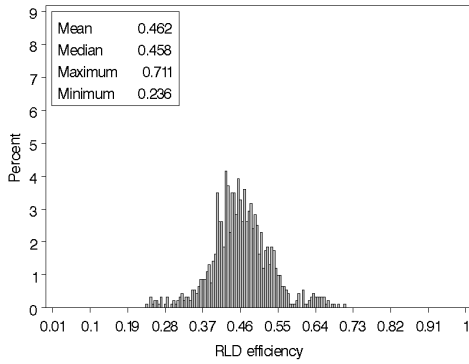
Figure 6: RLD -efficiencies for parameter space Ω_4 .



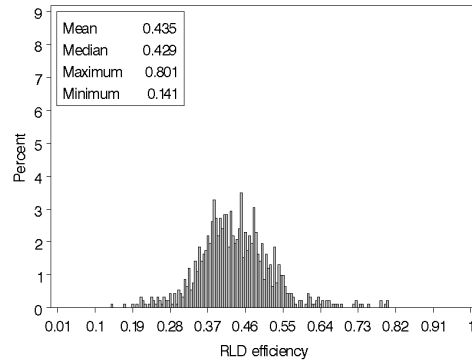
(a) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 1.51_8$



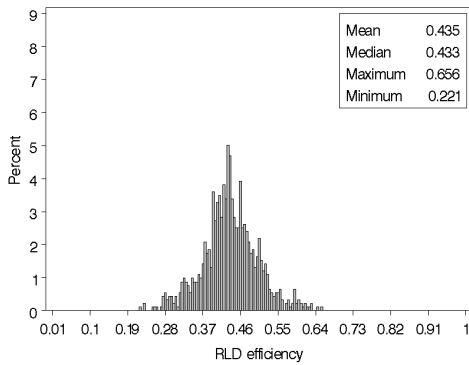
(b) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



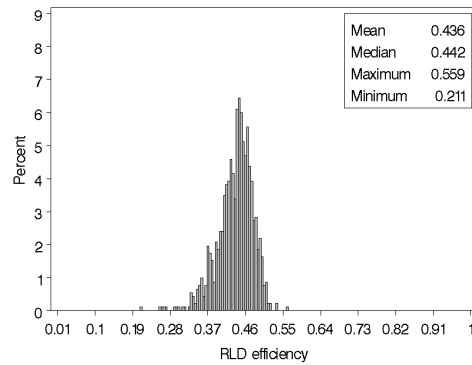
(c) Semi-Bayesian D -optimal mixed logit design with $\sigma_\beta = 0.51_8$



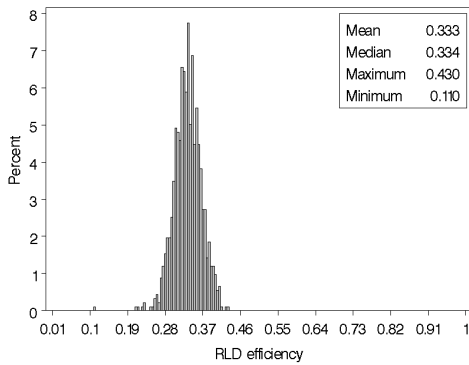
(d) Locally D -optimal mixed logit design with $\sigma_\beta = \mathbf{1}_8$



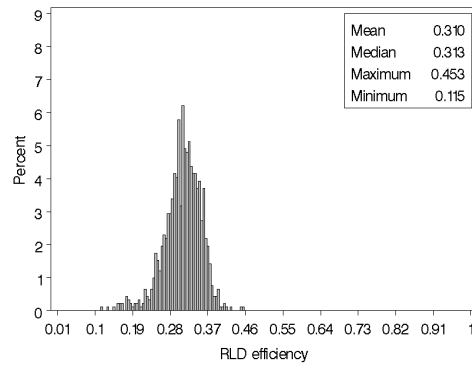
(e) Bayesian D -optimal multinomial logit design with covariance $9\mathbf{I}_8$



(f) Bayesian D -optimal multinomial logit design with covariance \mathbf{I}_8

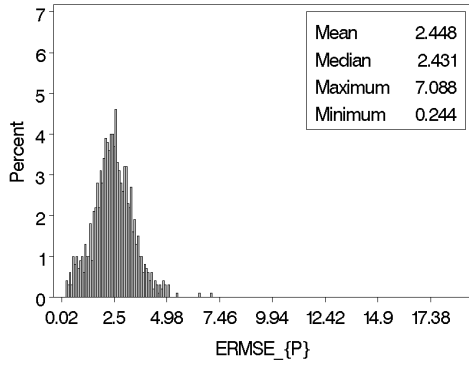


(g) Locally D -optimal multinomial logit design

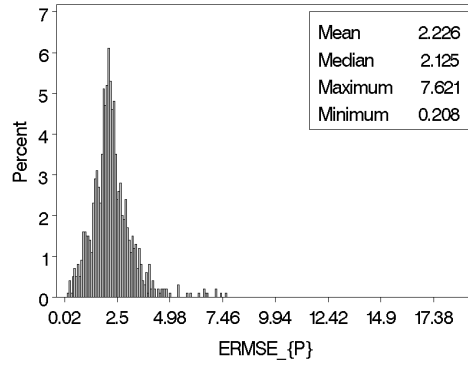


(h) Nearly orthogonal design

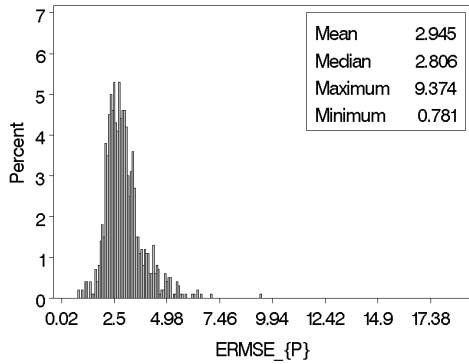
Figure 7: RLD -efficiencies for parameter space Ω_5 .



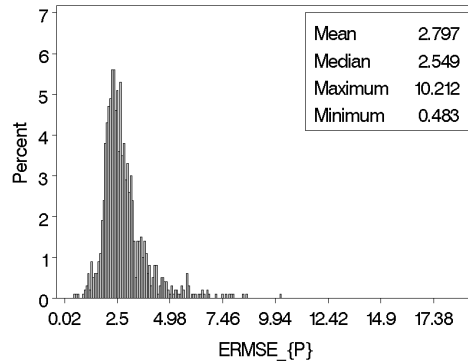
(a) Semi-Bayesian D -optimal mixed logit design with $\sigma_{\beta} = 1.51_8$



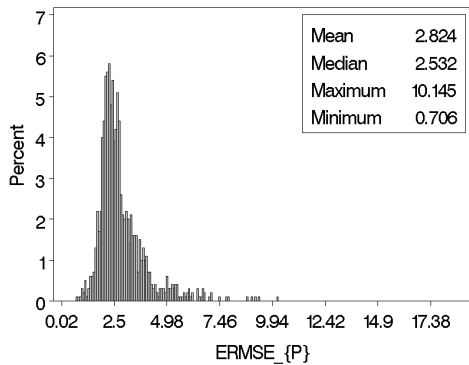
(b) Semi-Bayesian D -optimal mixed logit design with $\sigma_{\beta} = \mathbf{1}_8$



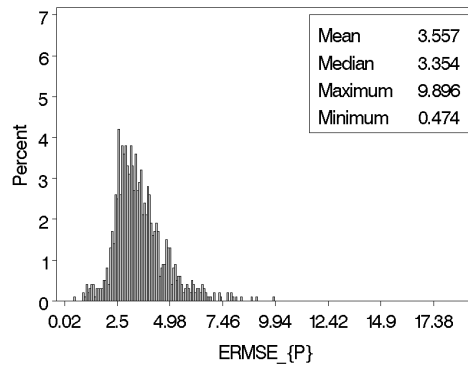
(c) Semi-Bayesian D -optimal mixed logit design with $\sigma_{\beta} = 0.51_8$



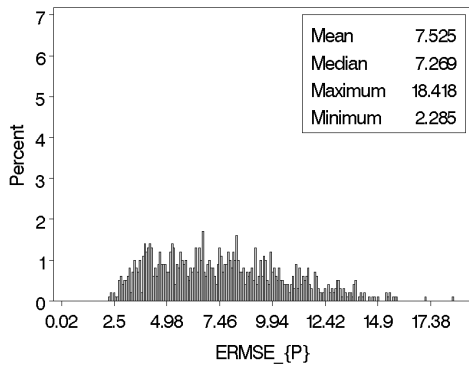
(d) Locally D -optimal mixed logit design with $\sigma_{\beta} = \mathbf{1}_8$



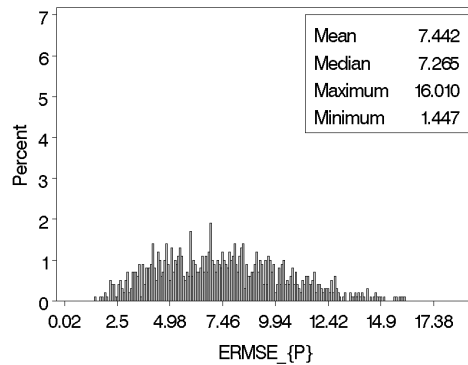
(e) Bayesian D -optimal multinomial logit design with covariance $9\mathbf{I}_8$



(f) Bayesian D -optimal multinomial logit design with covariance \mathbf{I}_8



(g) Locally D -optimal multinomial logit design



(h) Nearly orthogonal design

Figure 8: Expected root mean-squared prediction errors for Ω_2 .