Profile Construction in Experimental Choice Designs for Mixed Logit Models

Zsolt Sándor • Michel Wedel
Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
Department of Marketing and Marketing Research, University of Groningen, P.O. Box 800,
9700 AV Groningen, The Netherlands, and the University of Michigan Business School, 701 Tappan Street,
Ann Arbor, Michigan 48109
sandor@few.eur.nl • wedel@umich.edu

Abstract

A computationally attractive model for the analysis of conjoint choice experiments is the mixed multinomial logit model, a multinomial logit model in which it is assumed that the coefficients follow a (normal) distribution across subjects. This model offers the advantage over the standard multinomial logit model of accommodating heterogeneity in the coefficients of the choice model across subjects, a topic that has received considerable interest recently in the marketing literature. With the advent of such powerful models, the conjoint choice design deserves increased attention as well. Unfortunately, if one wants to apply the mixed logit model to the analysis of conjoint choice experiments, the problem arises that nothing is known about the efficiency of designs based on the standard logit for parameters of the mixed logit. The development of designs that are optimal for mixed logit models or other random effects models has not been previously addressed and is the topic of this paper.

The development of efficient designs requires the evaluation of the information matrix of the mixed multinomial logit model. We derive an expression for the information matrix for that purpose. The information matrix of the mixed logit model does not have closed form, since it involves integration over the distribution of the random coefficients. In evaluating it we approximate the integrals through repeated samples from the multivariate normal distribution of the coefficients. Since the information matrix is not a scalar we use the determinant scaled by its dimension as a measure of design efficiency. This enables us to apply heuristic search algorithms to explore the design space for highly efficient designs. We build on previously published heuristics based on relabeling, swapping, and cycling of the attribute levels in the design.

Designs with a base alternative are commonly used and considered to be important in conjoint choice analysis, since they provide a way to compare the utilities of profiles in different choice sets. A base alternative is a product profile that is included in all choice sets of a design. There are several types of base alternatives, examples being a so-called outside alternative or an alternative constructed from the attribute levels in the design itself. We extend our design construction procedures for mixed logit models to include designs with a base alternative and investigate and compare four design classes: designs with two alternatives, with two alternatives plus a base alternative, and designs with three and with four alternatives.

Our study provides compelling evidence that each of these mixed logit designs provide more efficient parameter estimates for the mixed logit model than their standard logit counterparts and yield higher predictive validity. As compared to designs with two alternatives, designs that include a base alternative are more robust to deviations from the parameter values assumed in the designs, while that robustness is even higher for designs with three and four alternatives, even if those have 33% and 50% less choice sets, respectively. Those designs yield higher efficiency and better predictive validity at lower burden to the respondent. It is noteworthy that our "best" choice designs, the 3- and 4-alternative designs, resulted not only in a substantial improvement in efficiency over the standard logit design but also in an expected predictive validity that is over 50% higher in most cases, a number that pales the increases in predictive validity achieved by refined model specifications.

(Conjoint Choice; Design Efficiency; Heterogeneity; Base-Alternative)

0732-2399/02/2104/0455/$05.00
1526-548X electronic ISSN

1. Introduction

Experimental choice analysis places the design of choice tasks under control of the researcher, offering the advantage of choosing a design that provides improved statistical properties of the model used to analyze it. Not all designs are equally useful, and several articles have addressed the problem of how to construct good ones (Lazari and Anderson 1994, Bunch et al. 1994, Huber and Zwerina 1996, Arora and Huber 2001, Sándor and Wedel 2001). The criterion mostly used by researchers in the field for what constitutes a good design is that it should provide as much information as possible on the parameters of interest. That is, it should lead to more efficient estimators.

Data from conjoint choice experiments are usually analyzed with multinomial logit models (Louviere and Woodworth 1983). An important drawback of that model is that does not accommodate consumer heterogeneity. Accounts of heterogeneity in the analysis of consumer behavior constitute an important topic in the recent marketing literature (Allenby et al. 1998, Allenby and Rossi 1999, Wedel et al. 1999). The ability to identify customer heterogeneity is important in new marketing approaches, such as one-to-one marketing, micro marketing, and mass customization. Several approaches to representing heterogeneity in choice data have been taken, the most important distinction between them being whether one assumes that the parameters of interest follow a continuous or a discrete distribution. The assumption of discrete mixing distribution leads to finite mixture multinomial logit models for conjoint choice experiments (Kamakura et al. 1994, DeSarbo et al. 1995). Such models connect elegantly to substantive theories of market segmentation, but the critique has been levied against them that they cannot adequately capture heterogeneity if the true underlying distribution is continuous (Allenby et al. 1998, Allenby and Rossi 1999). Many authors have turned attention to continuous mixing distributions, specifying random coefficient models. For example, Haaijer et al. (1998) proposed a random coefficient multinomial probit model for the analysis of conjoint choice experiments. Whereas their model is computationally demanding, a computationally more attractive—but conceptually similar—alternative is the mixed multinomial logit model (Revelt and Train 1998, McFadden and Train 2000): a logit model in which it is assumed that the coefficients follow a (normal) distribution across consumers. Such models have been applied to the analysis of scanner panel data (e.g., Allenby and Lenk 1994, 1995), but applications to conjoint choice experiments are less common. Recently, Huber and Train (2001) provided such an application and showed that the classical (Maximum Likelihood) and Bayesian (Gibbs sampling) approaches to estimate those models yield very similar results. With the advent of such powerful but highly parameterized models and sophisticated estimation methods, the quality of the design deserves attention, since efficiency of the estimators becomes more critical if many parameters are estimated. Unfortunately, in applying the mixed logit model to the analysis of conjoint choice experiments, nothing is known about the efficiency of classical designs for parameters of the mixed logit. The development of designs that are optimal for mixed logit models or other random effect models has not been addressed and constitutes the topic of this paper.

Conjoint choice experiments pose complicated problems of design construction. This is so because the design of experiments for the logit model, contrary to experimental design methods for linear models, requires knowledge of the values of its parameters. This situation arises because the amount of information on the parameters provided by the design is dependent on the value of those parameters. Several solutions to that problem have been proposed that include setting all parameters to zero, using ‘‘reasonable’’ parameter values, estimating parameter values from a pilot study, or obtaining them from judgments by consumers or managers (see Bunch et al. 1994, Kuhfeld et al. 1994, Lazari and Anderson 1994, Huber and Zwerina 1996, Arora and Huber 2001, Sándor and Wedel 2001).

We elaborate on Sándor and Wedel (2001), who present a method to provide more efficient designs based on prior information about the parameters and the associated uncertainty, elicited from managers. It is a Bayesian procedure that assumes a prior distribution of likely parameter values and optimizes the design across the entire range of the prior distribution. They propose a procedure to obtain prior information from managers that they illustrate in an empirical application, where they elicit prior information on parameter values as well as the associated uncertainty. However, neither this approach nor one of the currently available approaches for choice designs accommodates heterogeneity. Design construction is complicated if consumer heterogeneity is to be accommodated, as was shown by Lenk et al. (1996) for metric conjoint designs.

The work of Lenk et al. (1996) is a useful point of departure for our study. They proposed the application of a random coefficients (hierarchical Bayes) model to analyze metric conjoint data. Those models are designed to recover individual level estimates even from short conjoint questionnaires, using preference ratings and accommodating consumer heterogeneity. With the availability of such random coefficient models to analyze conjoint experiments, the issue of the design of those experiments became prevalent. Therefore, Lenk et al. (1996) develop experimental designs that provide high efficiency for parameters of their random coefficient model. Their approach of design and analysis, however, applies to standard metric conjoint experiments and not to choice experiments.

In this paper we address this issue and design choice experiments that provide more efficient estimates of parameters of the mixed logit model. We develop designs for the mixed logit, following the objectives of Lenk et al. (1996) and previous approaches for logit models, such as those proposed by Huber and Zwerina (1996), Sándor and Wedel (2001), and others. We provide an algorithm to construct the designs. The designs lead to large efficiency improvements, as we show in a Monte Carlo study. The next section describes the mixed logit model and proposes a design criterion for it. Subsequently, we describe the procedures that we use to generate the designs. In §3 we compare the performance of standard logit designs with that of mixed logit designs and investigate the sensitivity of the mixed logit designs to misspecification of the unknown parameters in design construction. In the Monte Carlo study we investigate the relative performance of designs in four different design classes (two, three, four alternative designs and two alternative designs with a base alternative). We conclude the paper and discuss topics for future research in §4.

2. Designs for the Mixed Logit

In the standard logit model, the utility for profile $j = 1, \ldots, J$ in choice set $s = 1, \ldots, S$ is specified as:

$$u_{js} = x_{js}'\beta + \varepsilon_{js}, \tag{1}$$

where x_{js} is a K-vector of the attribute levels of the alternative j, β is a K-parameter vector, and ε_{js} is an error term following an extreme value distribution. Assuming utility maximization, the probability that j is chosen from choice set s can be expressed in closed form:

$$p_{js} = \frac{\exp(x_{js}'\beta)}{\sum_{r=1}^{J} \exp(x_{rs}'\beta)}. \tag{2}$$

To account for heterogeneity in the parameters of the logit model across consumers, we assume that the parameter vector β is multivariate normally distributed with mean μ and variance Σ. Although covariance terms can be accommodated, we currently restrict attention to the case where $\Sigma^{1/2}$ is a diagonal matrix having $\sigma = (\sigma_1, \ldots, \sigma_K)'$ on its main diagonal. Hence, β can be written as $\beta = \mu + V\sigma$, where V is a diagonal matrix having the random vector $v = (v_1, \ldots, v_K)$ with pairwise independent standard normal elements on the diagonal. The model obtained is called the mixed or heterogeneous logit model (Brownstone and Train 1999). In the mixed logit, the probability of product j being chosen is

$$\pi_{js} = \int_{\mathbb{R}^K} p_{js}(v) \cdot \phi(v_1) \ldots \phi(v_K) dv, \tag{3}$$

where ϕ is the standard normal density function and

$$p_{js}(v) = \frac{\exp(x_{js}'[\mu + V\sigma])}{\sum_{r=1}^{J} \exp(x_{rs}'[\mu + V\sigma])}. \quad (4)$$

We are concerned with the construction of designs for the mixed multinomial logit model that provide high efficiency for the parameters β, and σ.[1] Our concern is to find a design defining the product profiles, given a number of choice sets, S, a number of alternatives in the choice sets, J, and a number of attribute level combinations, K. Following previous design construction procedures, we consider the sizes and numbers of choice sets as fixed by the researcher, prior to design construction. Our procedure can be designated as one for profile construction, given the other features of the choice design. Furthermore, in absence of a priori subject-specific information, subjects are considered exchangeable, and all receive the same choice sets. In addition, in some designs we allow each choice set to contain a base alternative, as detailed below.

The development of highly efficient mixed logit designs requires the evaluation of the information matrix of the mixed logit. Since we assume choices in different choice sets to be independent, the information matrix[2] takes the form

$$I(\mu, \sigma|X) = N \cdot \sum_{s=1}^{S} \begin{bmatrix} M_s'\Pi_s^{-1}M_s & M_s'\Pi_s^{-1}Q_s \\ Q_s'\Pi_s^{-1}M_s & Q_s'\Pi_s^{-1}Q_s \end{bmatrix}, \quad (5)$$

with

$$M_s = \int_{\mathbb{R}^K} [P_s(v) - p_s(v)p_s(v)']X \cdot \phi(v_1)\ldots\phi(v_K)dv$$

and

$$Q_s = \int_{\mathbb{R}^K} [P_s(v) - p_s(v)p_s(v)']XV \cdot \phi(v_1)\ldots\phi(v_K)dv,$$

where $p_s(v) = [p_{1s}(v), \ldots, p_{Js}(v)]'$, $P_s(v) = \text{diag}(p_{1s}(v), \ldots, p_{Js}(v))$, $X = [x_{11}, \ldots, x_{J1}, x_{12}, \ldots, x_{J2}, x_{1S}, \ldots, x_{JS}]'$, and $\Pi_s = \text{diag}(\pi_{1s}, \ldots, \pi_{Js})$.

The information in Equation (5) is not a scalar measure. A widely accepted one-dimensional measure of information is the determinant of the information matrix. It has been applied successfully to designs for logit models (Zacks 1977) and conjoint choice experiments (Bunch et al. 1994, Kuhfeld et al. 1994, Huber and Zwerina 1996). Researchers have more often used a transformed version of it, however, the D_P-error, which is based on the determinant of the covariance matrix scaled by its dimension K:

$$D_P\text{-error} = \det[I(\beta|X)]^{-1/K}. \quad (6)$$

We extend this criterion to the context of mixed logit designs, and as a measure of optimality for mixed logit designs we propose the D_M-error, which is the determinant of the inverse of the full information matrix to the power $1/2K$:

$$D_M\text{-error} = \det[I(\mu, \sigma|X)]^{-1/2K}. \quad (7)$$

This measure is directly optimized by our procedures. Whereas in the standard logit design only the parameter vector β needs to be fixed to determine designs with high efficiency (Huber and Zwerina 1996), we need to fix both $\mu = \mu_0$ and $\sigma = \sigma_0$, which may be considered a disadvantage of our and other

[1]The designs that we generate may not be strictly optimal for several reasons. First, a single scalar measure of the information is optimized. While the measure has intuitive appeal and has been frequently used in the literature, other measures are possible. Design "optimality" is only defined with respect to the particular scalar measure of information that we use. Second, as will be explained below, heuristic search measures are used to find the optimum. While those heuristics yield designs with improved efficiency, they may not yield "optimal" efficiency. Third, our design generating procedures rely on asymptotic approximations. Therefore, rather than referring to "optimal" designs, we will refer to designs with "improved efficiency," or "more efficient" designs.

[2]The information matrix, as shown in the Appendix, is derived in the usual way by computing the variance of the score function of the log-likelihood.

design generating procedures. We address this issue in detail later in the Monte Carlo study.

In the design-generating algorithm, the information matrix in (5) needs to be evaluated. Whereas in the standard logit for fixed β the information matrix has closed form, this is not the case for expression (5), since it involves integration over the distribution of the β's as shown in (5). We approximate the integrals by sampling from the multivariate normal distribution of $\beta \sim N(\mu, \Sigma)$. We use $R = 1{,}000$ repeated samples from that distribution. We need a relatively large number of draws, since for a small number of draws, chance fluctuations affect the estimates of the information matrix. We use procedures to generate draws that are more efficient than pseudorandom draws generated by the computer, based on orthogonal arrays (Tang 1993).

2.1. Algorithms for Design Construction

The designs we construct cannot be claimed to be strictly optimal. The reasons for that are, first, that the designs are generated under constraints (e.g., number of choice sets and number of alternatives per choice set) and, second, that finding the optimal design would necessitate that one search across the entire design space. Unfortunately, doing that is impossible, except for very small designs.[3] Whereas standard OR techniques could be used to explore the design space, we use heuristic procedures that have been developed specifically for generating choice designs. Thus, we use relabeling, swapping, and cycling, as proposed by Huber and Zwerina (1996) and Sándor and Wedel (2001). These heuristic procedures have been shown to perform quite well in constructing good designs by several authors (Huber and Zwerina 1996, Arora and Huber 2001, Sándor and Wedel 2001). The three stages operate as follows.

Relabeling permutes the levels of an attribute across choice sets and does that for all attributes and choice sets. Take Level 1 of Attribute 1 and exchange that with Level 2 of that attribute in all choice sets. Then take Level 1 of that attribute and exchange it with Level 3 across all choice sets, and so on. The relabeled design that arises is retained if it yields a lower D_M-error. All attributes are considered simultaneously.

Swapping involves switching two levels of the same attribute among alternatives within a choice set. Consider the first choice set. Take the level of Attribute 1 for the first alternative and swap that with the level of that attribute for the second alternative. The swapping algorithm verifies all those possible swaps starting with the first choice set and passing through all choice sets. If an improvement in information occurs, then the procedure returns to the first choice set and proceeds until no improvement is possible.

Cycling rotates the levels of the attributes cyclically. For example, if there are three levels, 1, 2, and 3, it replaces 1 by 2, 2 by 3, and 3 by 1. The algorithm starts with the levels of the first attribute in the first choice set and rotates cyclically all its levels until all possibilities are exhausted. (The number of subsequent rotations needed for this is equal to the number of levels minus one.) A cyclical rotation of the level of the first alternative is applied followed by subsequent cyclical rotations of all alternatives, again, until all possibilities are exhausted. After this, again, the algorithm rotates only the level of the first alternative and continues by rotating all levels afterwards, and so on until all possible cycles for that attribute are verified. The algorithm turns to the first attribute of the second choice set and further passes through all choice sets. Then it goes on to the next attribute until the last one. At each stage, if an improvement is made the procedure starts over from the first attribute in the first choice set. When no more improvement is possible, the procedure stops. We note that this cycling algorithm is different from that in Sándor and Wedel (2001) in that here the swaps are replaced by cycling the levels of the first alternatives.[4]

[3]For very small designs we experimented with a search of overall possible designs, which is possible, although time-consuming. The heuristic search algorithms did quite a good job in recovering the optimum. Obviously, there is no guarantee that this is also the case for larger designs.

[4]Gauss codes for the design generating procedures are available from the first author.

For more details of the swapping and relabeling design generating algorithms we refer to Huber and Zwerina (1996) and Sándor and Wedel (2001). Our design generating algorithms can be used for a wide range of designs. However, for designs within particular design classes, modifications to the algorithms make the search for better designs more effective. Whereas for designs with two alternatives per choice set the above procedures can be directly applied; below we present modifications to the algorithm for designs with a base-alternative and designs with three and four alternatives that make the search across the design space more effective.

We do not enforce criteria, such as maximal level balance, minimal level overlap, or orthogonality (Huber and Zwerina 1996) on the designs. Maximal level balance means that all levels of an attribute occur in frequencies that are as close as possible to each other. Minimal level overlap implies that within a choice set the number of times the same attribute level occurs should be as small as possible. (Information) orthogonality causes the parameter estimates to be independent, but this can be achieved only for one specific set of parameter values and is therefore not a useful criterion for the construction of mixed logit designs. Sacrificing these criteria, in particular minimal level overlap, allows us to generate statistically more efficient designs. However, since we start the algorithms with designs having that property, the design generating algorithms may conserve that property and the designs obtained may satisfy minimal level overlap.

2.2. Base Alternative Designs

A base alternative is a profile that is included in all choice sets of a design. Designs with a base alternative are commonly used, since they provide a way to compare the utilities of profiles in different choice sets. To our knowledge, no successful attempt has been made to develop procedures to generate such designs, while in 1996, Huber and Zwerina already stated this to be an important issue to be addressed in future research. In addition, currently nothing is known about the efficiency of base alternative designs relative to other conjoint choice designs. Thus, in this study we first generate base alternative designs with improved efficiency and then compare that efficiency to that of designs without a base alternative.

There are two important types of base alternatives. The first are the so-called outside alternatives (“I choose none of these,” “I retain the brand I currently own,” or “I do not make a purchase at this time from the alternatives indicated”), which do not depend upon any of the attributes included in the study. The second type of base alternative is constructed based on the attribute levels of interest but is constrained to be present in each choice set in the design. In the present study, we only deal with this type of attribute-based base alternative. It may be less common than the outside alternative in practice, but its construction is more challenging. Designs with outside alternatives can be constructed using our procedures by forcing the outside alternative to be in each choice set and applying the relabeling, swapping, and cycling algorithms to the nonoutside alternatives in each choice set.

Our method for base alternative designs with attribute-based base alternatives also uses swapping, relabeling, and cycling. We start with a design whose subdesign not containing the base alternative has the property of maximal level balance and minimal level overlap. Then we take that subdesign and optimize it in the same way as the designs without a base alternative described previously. Then we apply the swapping algorithm to the subdesign, relabeling to the base alternative, then cycling to the subdesign and relabeling again to the base alternative. We go on until no improvement of the objective function occurs. We note that the design criteria, level balance, minimal level overlap, and orthogonality cannot apply for designs with a base alternative.

2.3. Designs with Three and Four Alternatives

For constructing designs with 3- and 4-alternative choice sets we use the procedures relabeling, swapping, and cycling described previously. While relabeling can be used directly to these design classes, swapping and cycling need slight modifications in order to improve their effectiveness.

Swapping. We take the first choice set and apply swapping first in the usual manner (described above) to Alternatives 1 and 2 and then to Alternatives 2 and 3, subsequently, in the case of 3-alternative choice sets, and go on to Alternatives 3 and 4 in the case of 4-alternative choice sets. We go on to the next choice set, and so on, until no further improvement is possible.

Cycling. This procedure follows the same concept, namely, we start with the first attribute of the first choice set and apply cycling in the usual way, first to the first two alternatives and then to Alternatives 2 and 3, in the case of 3-alternative choice sets, and go on to Alternatives 3 and 4 in the case of 4-alternative choice sets. Then we move on to the first two alternatives of the second choice set, and so on, until no improvement is possible. We note that this procedure does not preserve the minimal level overlap property of the designs, and because of this it results in large potential improvements in the D_M-error of mixed logit designs.

3. Study on the Performance of Mixed Logit Designs

We investigate the performance of mixed logit designs. Since the designs are generated from parameter values that need to be fixed a priori, it is of particular interest to analyze their performance in situations when the parameter values assumed in generating the design deviate from the true ones. We are interested in two particular aspects of the relative performance of the designs: efficiency of the resulting parameter estimates and predictive ability. The corresponding measures are detailed in subsections.

The comparisons are all based on the following framework. We consider four design classes: designs of type $3^4/2/18$, that is, with 18 choice sets and in each set two alternatives with four attributes, each attribute having three levels; designs of the type $3^4/3/12$, with three alternatives in each of 12 choice sets; designs of the type $3^4/4/9$, with four alternatives in each of 9 choice sets and $3^4/(2 + 1)/12$ designs with three alternatives in each of 12 choice sets, where the third alternative in each choice set is a base alternative constructed on the basis of the attributes in the design and is included in each choice set.

We investigate the performance of the designs in a Monte Carlo study based on the following factors:

1. The mean of the coefficients assumed in the design are $\mu = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$ or $\mu = ½[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$.
2. The deviation of the assumed mean, μ, from the true (μ_0) mean is computed as $\mu_0 = \mu + \lambda v$, with μ as specified above, $v \sim N(0, I_K)$ and 31 values of $\lambda \in [0, 3]$.
3. The assumed standard deviation of the heterogeneity distribution is specified as $\sigma = [1.0 \ldots 1.0]'$, $\sigma = [0.5 \ldots 0.5]'$, or $\sigma = [0.2 \ldots 0.2]'$.
4. The true standard deviation of the heterogeneity distribution is taken to be $\sigma_0 = [1.0 \ldots 1.0]'$, $\sigma_0 = [0.5 \ldots 0.5]'$, or $\sigma_0 = [0.2 \ldots 0.2]'$.
5. The design type is either a 2-alternative, a 2-alternative + base, a 3-alternative, or a 4-alternative design.

This leads to a $2 \times 31 \times 3 \times 3 \times 4$ experimental design for the Monte Carlo study setup, resulting in 2,232 conditions.

In applying these design generating algorithms in practice, several procedures to choose the initial values of the parameters are available. These include estimating them from a pilot study or obtaining them from judgments by consumers or managers (see Huber and Zwerina 1996, Arora and Huber 2001, Sándor and Wedel 2001). Factors 1 and 2 above enable us to investigate the effect of misspecifying the mean of the mixed logit parameter. We use two sets of values of the means assumed in the design. These two values have a different scale, which makes it possible to study the effect of setting the scale of the means. For each of the two sets of assumed mean values we generate $\mu_0 = \mu + \lambda v$ as the true mixed logit mean parameter value, with $\lambda \in [0, 3]$ and $v \sim N(0, I_K)$. Hence, λ can be interpreted as the deviation of the true mean parameter value from the assumed mean parameter value. For a given value of λ, where we use 31 different values on [0, 3], we draw $r = 1, \ldots, R = 64$ mean parameter values μ_0^r as an orthogonal array-based Latin hypercube sample (Tang 1993). This enables us to generate a curve of the efficiency of the design against the degree of misspecification of the mean, increasing with λ, for each of the other conditions in the study.

The different values of the assumed and true heterogeneity parameters, varied in Factors 3 and 4, enable us to analyze the effect of misspecifying the heterogeneity in constructing the design. We use three levels of each of these factors, i.e., $\sigma_0 = [1.0 \ldots 1.0]'$, $\sigma_0 = [0.5 \ldots 0.5]'$, and $\sigma_0 = [0.2 \ldots 0.2]'$, and similar settings for σ. All nine combinations of those two factors are included in the study. Three of them pertain to situations where the variance is correctly specified in generating the design, the other six correspond to misspecified values. For a particular true value σ_0, comparing the curves for different specifications of σ allows one to appreciate the effect of misspecifying the heterogeneity, in generating the design, on its efficiency.

We investigate four different design types. The simplest design is the 2-alternative design. This design is a useful benchmark, although it tends to be used less often in conjoint choice experiments in practice, except when paired comparisons are needed. The 3- and 4-alternative designs are used much more often. The base alternative design is included to investigate the effect of adding a base alternative to the 2-alternative design. As described above, the base alternative itself is constructed from the attributes in the design.

The computations are done as follows. For each design class we construct a standard logit design, S, minimizing the D_P-error (6), and three mixed logit designs, M1, M2, M3, minimizing the D_M-error (7), as the objective function. The standard logit design, S, is constructed with the parameter values μ, the mixed logit designs are constructed using μ as the mean parameter and $\sigma = [1.0 \ldots 1.0]'$ for M1, $\sigma = [0.5 \ldots 0.5]'$ for M2, and $\sigma = [0.2 \ldots 0.2]'$ for M3 (see the description of factor 3 above). We construct the standard logit design with the same procedures as the mixed logit design, in that we use relabeling, swapping, and cycling. Each μ_0^r and σ are taken as the true mean and standard deviation parameter values and used in the evaluation of the information matrix and the two comparison measures. We measure the performance of the mixed logit designs by comparing them to the corresponding standard logit designs, and in addition we compare the base, 3- and 4-alternative designs to the 2-alternative design.

The designs obtained for the four design classes and $\mu = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$ are presented in Appendix 2 in Tables Al through A4. We note an interesting property of these designs regarding minimal level overlap, namely, that there is a positive correlation between the degree of level overlap and the heterogeneity parameter σ used in design construction. In order to make this statement more precise, we introduce an intuitive measure of level overlap for designs: The percentage of cases in which the columns of the choice sets satisfy the minimal level overlap criterion. For the 2- and 3-alternative designs, this measure amounts to calculating the percentage of columns in the choice sets whose elements are different, while for the 4-alternative designs it amounts to computing the percentage of columns that contain all three attribute levels, 1, 2, and 3. The results are contained in Table 1. The standard logit design has a very low-level overlap in all three design classes, while the M1 designs have the highest-level overlap. We can also notice that as σ drops to $[0.5 \ldots 0.5]'$ respectively $[0.2 \ldots 0.2]'$, the level overlap of the designs M2 and M3 decreases.

These findings seem to be in line with intuition. Since efficient standard logit designs maximize the information on the estimator of the parameters (which is, loosely speaking, the inverse of the "variability" in the model), these designs tend to maximize the variation in their attribute levels, leading to low-level overlap. This is the case because in the standard logit model, there is no other source of variation than the differences in attribute levels. The situation is different in the case of mixed logit designs, which maximize the information on the estimator of mixed logit parameters, including unobserved consumer heterogeneity through the value of σ_0. This case is another source of variation in addition to that provided by the attributes. This situation in turn implies that higher values of σ_0 allow for larger ‘‘variability’’ in the model, and consequently, attribute variation becomes less important (see Table 1). We conclude that relaxing the minimal overlap criterion in the construction of designs is more beneficial for the mixed logit designs than for the standard logit designs. However, this explanation is somewhat tentative, and we do not have a formal proof. In the next subsections we detail the computation of the performance measures in the Monte Carlo Study.

Table 1 Percentage of Level Overlap for Standard and Mixed Logit Designs in Different Design Classes

	*M*1	*M*2	*M*3	*S*
2-alternative	22	22	6	0
3-alternative	75	69	50	6
4-alternative	75	56	31	12

Note: *M*1, *M*2, and *M*3 are mixed logit designs constructed with $\sigma = [1.0 \ldots 1.0]'$, $\sigma = [0.5 \ldots 0.5]'$, and $\sigma = [0.2 \ldots 0.2]'$, respectively, while *S* is the standard logit design.

3.1. The D_M-error

For each value of μ and each σ, we compute the D_M-error $= \det[I(\mu, \sigma|D)]^{-1/2K}$, both for the standard ($D = S$) and the mixed logit ($D =$ M1, M2, M3) design. Then, the percentage difference of D_M-errors corresponding to these designs is computed and is averaged across the R draws of the true mean parameters. This measure can be interpreted as the increase in the number of respondents needed for the standard logit design to attain the same efficiency as the mixed logit design, when estimating a mixed logit model. This phenomenon is simple to show, because in the D_M-error the determinant of the information matrix is normalized by its dimension (Equation (7)), and since the information matrix is proportional to the number of subjects (Equation (5)), the D_M-error is too.

3.2. Expected Root Mean-Squared Prediction Error

Predictive validity has been of importance in the evaluation of conjoint models in practice. Although the procedures for generating mixed logit designs improve the efficiency of the estimates and not the predictive validity of the estimated models, improved efficiency will translate in better expected predictive validity as well. The criterion that we use for assessing the predictive validity of the designs is the expected root mean-squared error of the choices in holdout choice sets. To compute that measure, we proceed in the following manner. We construct a $3^4/2/6$ design for out-of-sample prediction (for all design classes in the simulation) and compute true and predicted probabilities by using the true and estimated parameters, respectively. Here we use the fact that the asymptotic distribution of the ‘‘estimated’’ parameters is known. We compute the Expected RMSE as:

$$ERMSE_D(n) = \int ([p(\hat{\beta}_D) - n]'[p(\hat{\beta}_D) - n])^{1/2} f(\hat{\beta}_D)\, d\hat{\beta}_D,$$

where $D = S$ or M1, M2, M3, for the standard logit and mixed logit designs, respectively. Here, n is the vector of choice frequencies in the holdout choice sets: The vector of probabilities computed for the true values μ_0^r, σ_0 of the parameters, $p(\hat{\beta}_D)$ is the corresponding vector of predicted probabilities computed for estimated parameter values $\hat{\beta}_D$, and $f(\hat{\beta}_D)$ is the asymptotic distribution of the estimates presented above. The expectation is again approximated by averaging over a large number of draws from the asymptotic distribution of the estimates. We compare the mixed logit design to the standard logit design by computing the percentage difference in the expected prediction RMSE between them. It may be observed that if the distribution of the estimates $f(\hat{\beta}_D)$ is more concentrated, as we expect to happen

Table 2 Predictive Validity Comparisons Based on Percentage Differences

	Deviation of True from Assumed Mean Parameters							
	0.5	1.5	0.5	1.5	0.5	1.5	0.5	1.5
	S and M1 Designs							
True Variance Parameters	2-Alternative		Base Alternative		3-Alternative		4-Alternative	
	$\mu = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$							
1.0	49.8	34.4	44.1	32.2	71.6	71.5	50.5	42.0
0.5	61.4	17.4	36.3	14.5	61.6	44.6	41.8	21.9
0.2	26.5	5.2	24.0	8.8	32.1	22.2	27.3	13.9
	$\mu = 1/2[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$							
1.0	28.6	35.9	51.9	46.7	65.0	66.0	56.1	43.9
0.5	38.7	22.1	40.1	16.9	62.9	49.3	48.6	34.4
0.2	18.1	12.3	16.5	8.4	38.4	31.7	36.0	26.8

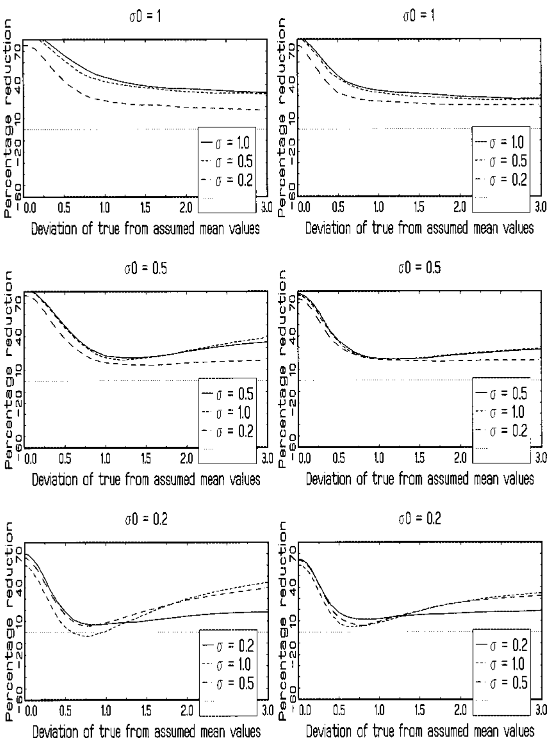
for the mixed logit design, the expected prediction error will be smaller.

3.3. Simulation Results

We used ANOVA to formally test the significance of the effects of the design factors and their interactions on the (logit transformed) percentage reduction in sample size relative to the standard logit design. The factors presented above lead to $2 \times 31 \times 3 \times 3 \times 4 = 2{,}232$ data points, which were analyzed for main effects and interactions of up to three factors. All main effects and most of the 2- and 3-factor interactions are significant. Therefore, it is not possible to present the results in a more condensed way by averaging over combinations of factors. However, to provide general insights, we present a few overall means, which should be interpreted with some care because of the presence of interactive effects. The percentage improvement for the two assumed mean values (Factor 1) are 42.4 and 45.9 for $\mu = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$ and $\mu = \frac{1}{2}[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$, respectively. These numbers represent the mean percentage reduction in sample size averaged over all other factors when mixed logit designs are used instead of standard logit designs. The mean percentage of improvement with respect to the design types (Factor 5) are 35.0, 46.5, 51.4, and 43.8 for 2-, base, 3-, and 4-alternative designs, respectively. Thus, the percentage reduction in sample size for mixed logit designs relative to standard logit designs is the highest for the 3-alternative designs (51.4%) and lowest for the 2-alternative designs (35%) averaged across all other factors in the study. Rather than providing the main effects, we describe the significant 2-factor interaction of assumed and true heterogeneity (Factors 3 and 4). If the true heterogeneity is characterized by $\sigma_0 = [1.0\ \ldots\ 1.0]'$, the mean percentage improvement under the assumed heterogeneity values $[1.0\ \ldots\ 1.0]'$, $[0.5\ \ldots\ 0.5]'$, $[0.2\ \ldots\ 0.2]'$ are, respectively, 59.6, 55.7, and 44.6; for $\sigma_0 = [0.5\ \ldots\ 0.5]'$ the mean percentages are 46.6, 48.7, and 41.8; and for $\sigma_0 = [0.2\ \ldots\ 0.2]'$ the corresponding mean percentages are 28.1, 36.9, and 35.4. Thus, the design constructed with an assumed heterogeneity equal to the true heterogeneity tends to perform best, which is as expected.

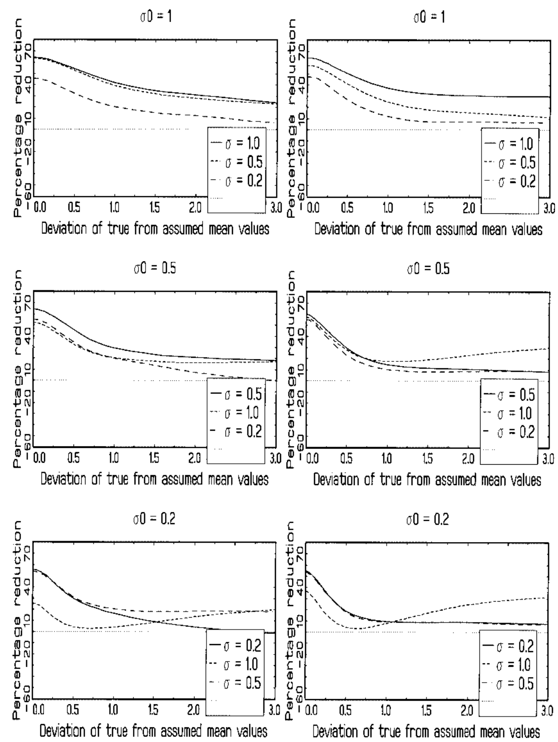
We provide a more detailed discussion of these results. Figures 1–4 present comparisons of standard and mixed logit designs from the same design class, while Figure 5 shows comparisons of mixed logit designs from different design classes. With the former comparisons we aim at analyzing the potential efficiency gains from using mixed logit designs instead

Figure 1 Graphs of the Percentage Reduction in Sample Size Needed with Respect to the Deviation of the Assumed from the True Mean Parameters



Note. These graphs are for 2-alternative mixed logit design relative to the standard logit design when the true heterogeneity variance is $\boldsymbol{\sigma}_0 = 1, 0.5, 0.2$ (top, middle and bottom panels. The designs are generated for two sets of means $\boldsymbol{\mu} = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ and $\boldsymbol{\mu} = 1/2[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ (left versus right panels) and three values of the assumed heterogeneity variance $\boldsymbol{\sigma} = 1, 0.5, 0.2$ (three lines in the graphs).

Figure 2 Graphs of the Percentage Reduction in Sample Size Needed with Respect to the Deviation of the Assumed from the True Mean Parameters

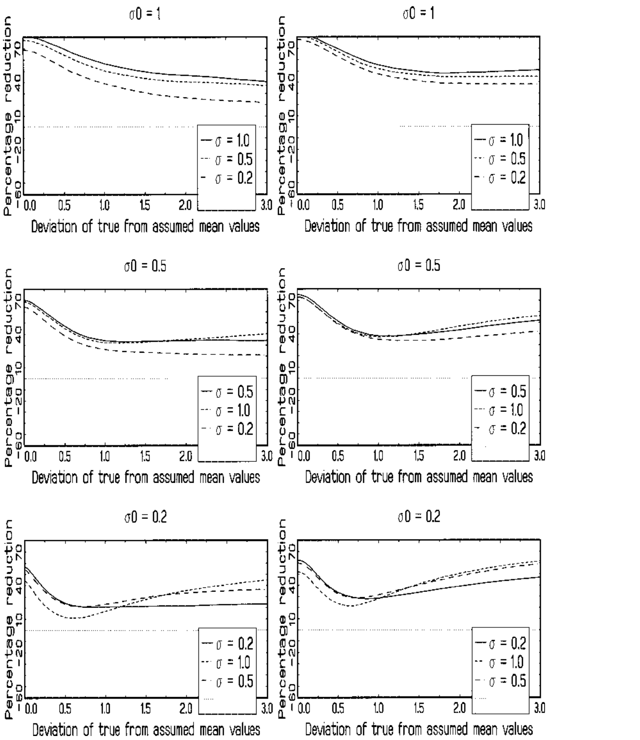


Note. These graphs are for base alternative mixed logit design relative to the standard logit design when the true heterogeneity variance is $\boldsymbol{\sigma}_0 = 1, 0.5, 0.2$ (top, middle and bottom panels). The designs are generated for two sets of means $\boldsymbol{\mu} = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ and $\boldsymbol{\mu} = 1/2[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ (left versus right panels) and three values of the assumed heterogeneity variance $\boldsymbol{\sigma} = 1, 0.5, 0.2$ (three lines in the graphs).

of standard logit designs when the true mean and heterogeneity parameters of the model are unknown. These figures also show the effects on design efficiency when the parameters are misspecified. For analyzing the effect of misspecifying the heterogeneity parameters the standard logit design serves as a useful benchmark since it does not depend on these parameters. With the graphs in Figure 5 we aim at directly comparing different mixed logit designs in cases of parameter misspecification.

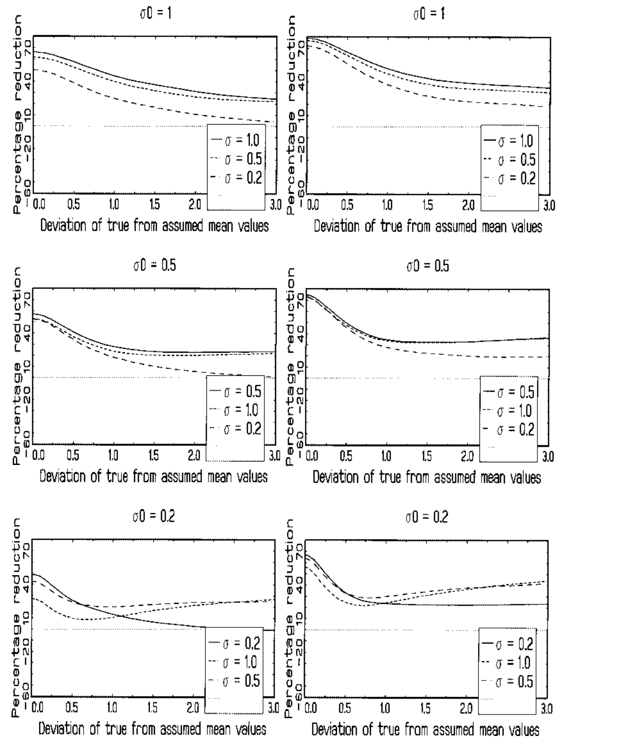
Figures 1–4 show the efficiency gains of mixed logit designs with respect to standard logit designs for the four design classes: the 2-alternative design, the base-alternative design, the 3-alternative design, and the 4-alternative design, respectively. Each graph contains six panels grouped in two columns, where the left columns show the results for an assumed value of $\mu = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$ and the right columns show those for $\mu = ½[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$. The three panels in each column correspond to the three different true values of the heterogeneity variance: $\sigma_0 = [1.0 \ldots 1.0]'$ (top panel), $\sigma_0 = [0.5 \ldots 0.5]'$ (middle panel), or $\sigma_0 = [0.2 \ldots 0.2]'$ (bottom panel). In each of these graphs, the

Figure 3 Graphs of the Percentage Reduction in Sample Size Needed with Respect to the Deviation of the Assumed from the True Mean Parameters



Note. These graphs are for 3-alternative mixed logit design relative to the standard logit design when the true heterogeneity variance is $\boldsymbol{\sigma}_0 = 1, 0.5, 0.2$ (top, middle and bottom panels). The designs are generated for two sets of means $\boldsymbol{\mu} = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ and $\boldsymbol{\mu} = 1/2[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ (left versus right panels) and three values of the assumed heterogeneity variance $\boldsymbol{\sigma} = 1, 0.5, 0.2$ (three lines in the graphs).

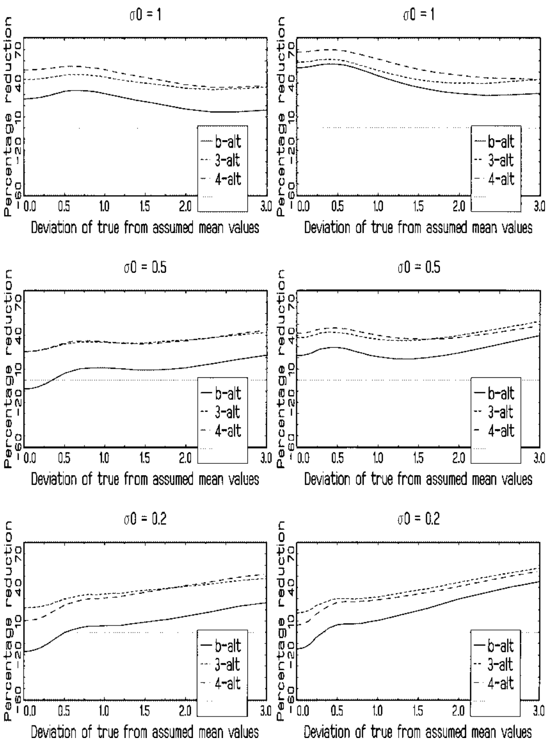
Figure 4 Graphs of the Percentage Reduction in Sample Size Needed with Respect to the Deviation of the Assumed from the True Mean Parameters



Note. These graphs are for 4-alternative mixed logit design relative to the standard logit design when the true heterogeneity variance is $\boldsymbol{\sigma}_0 = 1, 0.5, 0.2$ (top, middle and bottom panels). The designs are generated for two sets of means $\boldsymbol{\mu} = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ and $\boldsymbol{\mu} = 1/2[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ (left versus right panels) and three values of the assumed heterogeneity variance $\boldsymbol{\sigma} = 1, 0.5, 0.2$ (three lines in the graphs).

D_M-error is expressed relative to the D_M-error of the corresponding standard logit design, so that the vertical axes can be interpreted as the percentage reduction in the number of subjects needed to achieve the same efficiency as the standard logit design. The horizontal axis of each graph presents the deviation of the true from the assumed values of the mean parameters, previously denoted by λ, which takes on 31 grid values in the interval [0, 3] as described above (for each λ we draw 64 true parameter values, and for each of them we compute the percentage difference of the D_M-errors of the standard and mixed logit designs and average these percentage differences over the 64 draws). The larger the deviation λ, the larger the percentage of the true values that are far from the ones assumed in constructing the design, so that the degree of misspecification of the mean parameters of the mixed logit model increases from left to right in each graph. In each panel there are three curves that correspond to different values of σ assumed in generating the designs and computing the D_M-error, respectively: $\sigma = [1.0 \ldots 1.0]'$, $\sigma = [0.5 \ldots 0.5]'$, and $\sigma = [0.2 \ldots 0.2]'$. The differences between those three lines in each panel

Figure 5 Graphs of the Percentage Reduction in Sample Size Needed with Respect to the Deviation of the Assumed from the True Mean Parameters



Note. These graphs are for base, 3- and 4-alternative mixed logit design relative to the 2-alternative mixed logit design (the three lines in the graphs) when the true heterogeneity variance is $\boldsymbol{\sigma}_0 = 1$, 0.5, 0.2 (top, middle and bottom panels). The designs are generated for two sets of means $\boldsymbol{\mu} = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ and $\boldsymbol{\mu} = 1/2[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]$ (left versus right panels) and heterogeneity variance $\boldsymbol{\sigma} = 1$.

enable one to identify the effect of a misspecified covariance matrix in design construction. For example, if $\sigma_0 = [1.0 \ldots 1.0]'$, and for the line corresponding to $\sigma = [1.0 \ldots 1.0]'$ in one of the top panels of the figures, the assumption of the covariance matrix in design construction is correct, but the other two lines present situations of misspecification.

The first impression when looking at these figures is that although they are similar qualitatively, they are different quantitatively, both for different μ's and different design classes. There are many features that are intuitive and occur across all the graphs. The most prominent is that all curves are above 0, which means that all mixed logit designs are more efficient than the corresponding standard logit designs and are thus not only better on average, as the main effects mentioned above show, but uniformly better in all conditions in the study. The efficiency improvements are, on average, around 20–30% for "average" misspecification and rise to more than 50% if the assumed mean parameter values are fairly close to the true ones. This implies that with mixed logit designs a reduction in the number of respondents of 20 up to 50% may yield the same efficiency as that of standard logit designs.

Another typical feature is that in the top panels, that is, for true heterogeneity parameters $\sigma_0 = [1.0 \ldots 1.0]'$ the M1 designs are the best, followed by the M2 designs. This confirms the interaction effect described above and is expected because the M1 designs are constructed with the variance parameters $\sigma = [1.0 \ldots 1.0]'$, equal to the true ones. We notice that the difference in efficiency between M1 and M2 is generally quite small, on average approximately 5%. The difference between M2 and M3 is higher: It can go up to approximately 20%, which corroborates the results provided above. Thus, these mixed logit designs are fairly insensitive to mild misspecification of the heterogeneity parameters. The designs constructed with the true heterogeneity parameters and mean parameters close to the true ones are more efficient than those with misspecified heterogeneity parameters, which is seen in the top, middle, and bottom panels, reflecting different levels of true heterogeneity. However, the situation is different if the mean parameters deviate more from the true ones since the M1 designs tend to become the most efficient in the right half of both the middle and bottom (less true heterogeneity) panels, and the M2 designs become more efficient than the M3 designs in the bottom panels. The M3 designs dominate in the bottom panels for low deviation of the mean values, where the true heterogeneity is equal to variance used in their construction. This phenomenon is obviously a minor part of all possible cases. We believe it is intuitive if we interpret the heterogeneity parameter assumed for constructing the designs as a tool that can account for possible misspecification of the true mean parameters. In this regard, large heterogeneity parameters are expected to produce designs that are more robust to misspecification of the mean parameters than small heterogeneity parameters. Taking into account all possible types of misspecification we conclude that it is better to assume larger heterogeneity parameters in design construction since these parameters lead to more efficient mixed logit designs, especially if one is uncertain about the true mean values.

The mixed logit designs constructed with a smaller scale of μ, that is, $\mu = \frac{1}{2}[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$, tend to have a somewhat higher efficiency relative to the corresponding standard logit designs. Differences are not large, as can also be seen from the means presented above, but are here seen to accrue especially in cases where the assumed standard deviation is $\sigma = [1.0 \ldots 1.0]'$ and when the true means are relatively far from the assumed ones.

Similar to Figures 1–4, Figure 5 contains six panels where the two columns correspond to the two different assumed mean values μ, and the top, middle, and bottom panels correspond to the three true heterogeneity parameters σ_0. The curves represent comparisons of the base, 3-, and 4-alternative mixed logit designs to the 2-alternative mixed logit design where all four designs are constructed with the heterogeneity parameter $\sigma = [1.0 \ldots 1.0]'$. What is apparent from all graphs is first, that apart from a minor part of deviations between 0 and 0.5 in three of the panels, the base, 3-, and 4-alternative designs are more efficient than the 2-alternative design for all three values of true heterogeneity (misspecification) and, second, that the base alternative design is significantly less efficient than the 3- and 4-alternative designs. In the top panels, where the true heterogeneity parameters are the same as the assumed ones, the 4-alternative designs are more efficient than the 3-alternative designs. In the middle panels, where the heterogeneity is moderately misspecified, these two types of designs have approximately the same performance. In the bottom panel, where the difference between the assumed and true heterogeneity is higher, the 3-alternative designs appear to be slightly more efficient than the 4-alternative designs. Overall, based on the D_M-error comparisons from all six graphs the 4-alternative designs appear to be slightly more efficient than the 3-alternative designs. Thus, based on these findings we draw the preliminary conclusion that increasing the number of alternatives in a choice set improves efficiency of the design, except when the heterogeneity of the mixed logit is strongly misspecified. The differences between the left and right panels, representing different values of μ, are small.

Finally, we are interested in whether the designs with improved efficiency also have better predictive validity. Table 2 shows the expected predictive validity of the 2-, base, 3-, and 4-alternative standard and mixed logit designs. The table contains two panels of results that show comparisons of expected prediction RMSE of standard logit designs to M1 designs with 2-, base, 3-, and 4-alternatives. Similar to the D_M-error comparisons, the reported results here are average percentage differences of expected prediction RMSEs. The first panel represents the case of $\mu = [-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$ while the third panel represents the case $\mu = \frac{1}{2}[-1\ 0\ -1\ 0\ -1\ 0\ -1\ 0]'$. The predictive validity is evaluated for various values of the assumed σ, as shown in the rows of the table. We evaluate these designs for two values of the deviation of the true from the assumed means: 0.5 and 1.5.[5]

ANOVA revealed significant main effects of the deviation of the assumed mean (Factor 2; mean percentages 42.8 and 30.1 corresponding to deviation values 0.5 and 1.5, respectively), the true heterogeneity (Factor 4; mean percentages 49.4, 38.3, and 21.8 for the true heterogeneity values $[1.0 \ldots 1.0]'$, $[0.5 \ldots 0.5]'$, $[0.2 \ldots 0.2]'$, respectively), and the design type (Factor 5; mean percentages 28.4, 29.2, 51.4, and 36.9 for 2-, base, 3- and 4-alternative designs, respectively). The 2-factor interactions of these data were also significant.

In terms of the designs compared, the results from Table 2 correspond to the curves of the M1 and

[5]The Monte Carlo study for predictive validity is somewhat less elaborate than those for the other two measures above, because of the very large computation times required to approximate the integrals through simulation at each of the design points.

standard logit designs in Tables A1–A4. The results for the expected RMSE of prediction are somewhat similar to the result reported in Figures 1–4. The results show the overwhelming superiority of mixed logit design over standard logit designs in terms of predictive validity, if there is heterogeneity. For large misspecification of the—mean and variance—parameters, that is, for deviation of mean parameters equal to 1.5 and true variances equal to 0.2, the percentage differences are well above 5%. The average percentage difference is about 30–40%. According to these results, the largest difference between the standard logit and M1 design is obtained in the class of 3-alternative designs (minimum 22.2%, maximum 71.6% improvement).

4. Conclusion

This paper provides compelling evidence that mixed logit designs provide more efficient parameter estimates for the mixed logit model than standard logit designs and substantially higher predictive validity. We find our results in that respect striking. We summarize here three of our main findings. First, in the cases investigated, designs that include a base alternative are more robust to deviations from the parameter values assumed in the designs, while that robustness is even higher for designs with 3- and 4-alternatives, even given that those have 33%, respectively 50% less choice sets. Thus, those designs yield higher efficiency and better predictive validity at lower burden to the respondent. Secondly, the improvements tend to be larger if the assumed heterogeneity parameters in the design construction are larger. Third, our "best" choice design resulted not only in a substantial improvement in efficiency over the standard logit design but also in an improved expected predictive validity of around 65%, a number that pales the increases in predictive validity achieved by model refinements. Based on these results we have the following recommendations.

- If the mean and variance parameters of the mixed logit are known fairly precisely, for example, obtained from estimates of a prior or pilot study, then the design should be generated with those values.
- The design class should contain a larger number of choice alternatives (for example, 3 or 4) per choice set. Here 4-alternative designs do slightly better than 3-alternative designs and clearly better than 2-alternative and base alternative designs. Designs from these latter two classes cannot be recommended in general for estimating the mixed logit. There may be substantive reasons to use designs with base alternative.
- If the parameter values are unknown, a heuristic procedure for generating a design with improved efficiency and predictive validity is to choose a zero mean and unit variances to generate the design. These parameter settings seem to present a useful default for practical applications.

While we believe that we have presented significant advance in generating designs for conjoint choice experiments, there are also limitations. In our procedure, next to having to predetermine the values of the mean regression parameters as in a standard logit design, we also have to choose values for the variances of the regression parameters. However, our study shows that, with the exception of the case where the true variances are very small, the efficiency of the design is not much affected if the mean and/or variance parameters are misspecified in constructing the design. Nevertheless, the procedures applied in this study may be extended to Bayesian designs based on prior values of the mean and variance parameters analogous to Sándor and Wedel (2001), but currently the computational time required for such design generating procedures seems prohibitive.

An important topic for future research is to extend our procedures to enable the inclusion of the number of choice sets and the number of alternatives within a choice set in design generation. From both a statistical and measurement perspective, increasing the size of the choice set and simultaneously decreasing the number of choice sets is advisable. However, the conclusion is limited to choice sets with 2-, 3-, and 4-alternatives investigated in this study. Therefore, we believe more work is needed to further explore the trade-off of choice set size and number. Another issue of interest is to investigate the design-generating algorithms in the situation where a nondiagonal covariance matrix is to be estimated in the mixed logit model. The algorithms themselves are a topic for future study. While these heuristic procedures seem to work well for the construction of choice designs to which they have been tailored, the performance of improved search procedures, for example, integer programming (Goldengorin and Sierksma 1999) and genetic algorithms (Hamada et al. 2001), should be investigated.

Acknowledgments
The authors acknowledge the very useful contributions of Peter Kooreman and Tom Wansbeek in various stages of this project. In addition, the authors are much indebted to the editor, an area editor, and two anonymous reviewers for their useful comments and suggestions. Part of this study was carried out while Z. Sándor was a Ph.D. candidate at the University of Groningen.

Appendix 1. Information Matrix for the Mixed Logit Model

In this Appendix we derive briefly the information matrix for the mixed logit model given in (5). We use the notation from Equations (1), (2), (3), and (4). For saving notation, integrals like $\int_{\mathbb{R}^K} M(\cdot)\phi(v_1)\ldots\phi(v_K)\, dv$ are written as $\int(\cdot)\, d\Phi$ (e.g., $\int_{\mathbb{R}^K} p_j(v)\phi(v_1)\ldots\phi(v_K)dv \equiv \int p_j d\Phi$). Assume for the moment that there is only one choice set, i.e., $S = 1$, and suppress the notation in the subscript corresponding to it.

The log-likelihood function of this model is

$$L = N \cdot \sum_{j=1}^{J} f_j \cdot \ln \pi_j,$$

where f_j denotes the observed number of purchases of product j divided by the total number of purchases, N. For maximum likelihood estimation we are interested in the first- and second-order derivatives of the log-likelihood:

$$\frac{\partial L}{\partial \mu} = N \cdot \sum_{j=1}^{J} \frac{f_j}{\pi_j} \cdot \frac{\partial \pi_j}{\partial \mu} \quad \text{and} \quad \frac{\partial L}{\partial \sigma} = N \cdot \sum_{j=1}^{J} \frac{f_j}{\pi_j} \cdot \frac{\partial \pi_j}{\partial \sigma}.$$

Hence, we need the following formulae:

$$\frac{\partial \pi_j}{\partial \mu} = \int \frac{\partial p_j}{\partial \mu}\, d\Phi, \qquad \frac{\partial \pi_j}{\partial \sigma} = \int \frac{\partial p_j}{\partial \sigma}\, d\Phi,$$

$$\frac{\partial p_j}{\partial \mu} = (x_j - X'p)p_j, \qquad \frac{\partial p_j}{\partial \sigma} = V(x_j - X'p)p_j.$$

Combining these, we obtain

$$\frac{\partial \pi_j}{\partial \mu} = x_j\pi_j - X' \int pp_j\, d\Phi \quad \text{and} \quad \frac{\partial \pi_j}{\partial \sigma} = \int Vx_jp_j - VX'pp_j\, d\Phi.$$

Now we are able to express the log-likelihood as:

$$\frac{\partial L}{\partial \mu} = N \cdot \sum_{j=1}^{J} \frac{f_j}{\pi_j} \cdot \left(x_j\pi_j - X' \int pp_j\, d\Phi \right)$$
$$= N \cdot \left(X'f - X' \left[\int pp'\, d\Phi \right] \Pi^{-1} f \right) = N \cdot X' \int (P - pp')\, d\Phi \cdot \Pi^{-1} f$$

and

$$\frac{\partial L}{\partial \sigma} = N \cdot \sum_{j=1}^{J} \frac{f_j}{\pi_j} \int (Vx_jp_j - VX'pp_j)\, d\Phi$$
$$= N \cdot \int VX'(P - pp')\, d\Phi \cdot \Pi^{-1} f, \tag{8}$$

where $f = (f_1, \ldots, f_J)'$, $P = \text{diag}(p_1, \ldots, p_J)$, and $\Pi = \int P d\Phi$.

The information matrix is defined as:

$$I(\mu, \sigma|X) = \begin{bmatrix} E\frac{\partial L}{\partial \mu}\frac{\partial L}{\partial \mu'} & E\frac{\partial L}{\partial \mu}\frac{\partial L}{\partial \sigma'} \\ E\frac{\partial L}{\partial \sigma}\frac{\partial L}{\partial \mu'} & E\frac{\partial L}{\partial \sigma}\frac{\partial L}{\partial \sigma'} \end{bmatrix}.$$

For simplifying this expression we use the fact that f is assumed to have the multinomial distribution with $Ef = \pi$. Then f has the property that $Eff' = (1/N)\Pi + (1 - 1/N)\pi\pi'$. Using these we finally obtain (5) for $S = 1$. If the design has more than one choice set, since we assume that choices in different choice sets are made independently, the log-likelihood function will be the sum of the log-likelihoods corresponding to the choice sets. This leads to Formula (5).

Note that the standard logit model (Equation (2)) arises as a special case of the mixed logit model (Equation (4)) for $\sigma \downarrow 0$. The score function for the logit similarly arises from that of the mixed logit in the limit since $\partial L/\partial \sigma \to 0$ (see (8)) because the expectation $E[VX'] = 0$.

Appendix 2. Tables with Designs in the Four Design Classes

Table A1 Designs with Two Alternatives

Choice Set	Profile	S: Standard Logit				M1: Mixed Logit				M2: Mixed Logit				M3: Mixed Logit			
		Attributes				Attributes				Attributes				Attributes			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	I	1	3	3	2	3	3	1	1	3	2	3	3	3	3	1	1
	II	3	2	1	3	3	1	1	1	3	2	1	2	1	2	3	3
2	I	1	2	2	2	2	3	3	3	1	1	3	3	2	1	3	1
	II	2	1	1	3	1	3	1	3	2	3	3	2	1	3	2	3
3	I	3	2	1	2	3	3	2	2	1	3	1	3	1	3	2	1
	II	2	1	3	1	2	3	1	2	3	3	3	1	3	2	3	1
4	I	1	2	1	3	1	2	3	2	1	2	3	2	2	3	3	3
	II	2	1	2	2	3	3	3	2	3	3	3	3	3	3	2	1
5	I	3	1	2	3	3	2	3	2	2	2	2	2	2	2	2	3
	II	2	3	3	2	2	3	2	1	3	2	3	3	2	1	2	1
6	I	2	2	2	1	1	1	3	3	3	3	3	3	1	3	3	2
	II	1	3	1	2	2	3	3	2	3	3	3	1	3	1	2	1
7	I	3	1	2	1	3	1	1	1	1	3	1	2	1	3	2	3
	II	2	2	1	3	1	2	2	2	2	2	2	1	3	2	3	2
8	I	2	3	1	3	3	2	2	3	1	3	1	1	2	2	1	1
	II	3	1	3	1	1	1	2	1	3	1	2	2	2	1	3	2
9	I	2	2	3	1	1	1	2	2	3	3	1	3	2	3	3	1
	II	1	3	1	2	2	2	1	3	2	1	3	1	3	1	1	2
10	I	1	3	2	3	2	2	2	2	1	3	2	1	3	2	1	3
	II	3	1	3	2	1	1	3	1	2	2	1	2	1	2	3	1
11	I	3	1	2	3	2	2	2	3	1	2	3	1	1	2	2	2
	II	2	2	3	1	2	2	2	1	3	2	2	2	2	3	1	3
12	I	3	3	3	1	1	1	2	3	1	3	3	1	2	3	2	2
	II	2	2	2	3	3	3	3	1	1	1	2	1	3	1	3	3
13	I	3	1	1	2	1	2	1	1	1	2	2	3	2	2	3	1
	II	1	3	2	1	1	3	3	2	1	1	3	1	3	3	2	2
14	I	2	3	1	1	1	3	1	2	2	3	3	2	2	1	3	1
	II	1	2	3	2	3	1	2	3	2	2	1	2	3	2	1	2
15	I	3	2	3	1	2	1	2	2	2	1	3	3	1	3	1	1
	II	2	3	2	2	1	3	3	3	3	2	1	1	2	1	2	2
16	I	2	2	1	2	1	2	2	1	3	2	3	2	1	2	1	3
	II	1	1	3	3	1	1	1	2	2	3	1	3	2	3	3	1
17	I	1	1	3	3	3	2	1	1	1	2	3	2	3	2	2	2
	II	3	3	2	1	2	3	2	3	3	1	1	1	1	2	1	3
18	I	2	1	2	3	2	1	3	1	3	2	2	1	2	2	2	3
	II	3	3	1	1	3	3	1	3	2	1	2	1	1	1	3	2

Table A2 Designs with a Base Alternative

Choice Set	Profile	S: Standard Logit Attributes				M1: Mixed Logit Attributes				M2: Mixed Logit Attributes				M3: Mixed Logit Attributes			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	I	3	3	2	1	3	3	3	1	2	2	3	2	3	3	2	2
	II	2	2	3	2	3	1	3	1	2	2	2	1	2	3	2	1
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
2	I	3	2	1	2	3	1	3	2	2	1	2	2	1	3	1	3
	II	2	1	3	3	3	2	3	3	2	1	3	3	3	1	3	3
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
3	I	3	3	1	1	1	2	2	1	3	1	1	1	3	3	1	1
	II	1	2	2	3	3	2	2	2	1	2	2	2	3	2	2	3
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
4	I	2	3	3	2	3	3	3	2	1	3	3	3	1	3	3	3
	II	3	1	2	3	3	3	3	3	1	1	3	3	3	1	2	2
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
5	I	1	2	3	2	1	2	2	1	3	3	3	1	2	3	2	1
	II	2	3	1	1	1	2	3	2	2	2	1	2	3	2	3	1
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
6	I	3	3	2	2	3	1	1	3	1	3	1	3	3	1	1	3
	II	2	2	3	3	3	3	1	2	3	3	2	2	3	3	1	2
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
7	I	3	1	3	2	1	3	2	3	3	2	3	2	1	1	2	3
	II	2	2	2	3	2	2	2	2	2	2	3	1	3	1	1	2
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
8	I	2	2	1	1	3	2	2	3	3	1	3	3	2	3	3	1
	II	1	1	3	1	2	2	1	3	3	1	1	3	1	3	2	1
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
9	I	1	2	1	2	3	1	3	2	2	3	1	2	3	2	1	3
	II	1	1	3	1	3	1	1	1	3	2	1	1	2	1	3	3
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
10	I	3	1	1	3	2	2	3	3	2	3	3	1	2	3	1	3
	II	1	3	3	1	2	3	3	1	1	1	2	3	3	3	3	1
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
11	I	3	2	2	1	2	1	2	1	3	2	1	2	2	2	2	1
	II	1	3	1	3	1	1	1	3	3	3	1	3	1	3	1	3
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2
12	I	3	2	3	1	3	1	3	1	3	1	2	2	3	1	2	1
	II	1	3	2	3	3	1	3	2	2	2	2	3	2	3	1	3
	III	2	1	2	2	3	2	3	1	1	3	3	1	1	2	3	2

Table A3 Designs with Three Alternatives

Choice Set	Profile	S: Standard Logit				M1: Mixed Logit				M2: Mixed Logit				M3: Mixed Logit			
		Attributes				Attributes				Attributes				Attributes			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	I	1	3	3	1	3	1	2	2	2	1	2	2	3	1	3	2
	II	3	2	1	2	3	2	1	1	1	1	3	1	2	1	1	3
	III	2	1	2	3	1	1	1	2	3	2	2	1	1	3	2	1
2	I	3	1	2	1	2	1	2	1	1	2	2	2	3	1	3	1
	II	1	2	3	2	2	1	1	2	1	1	3	3	3	3	1	3
	III	2	3	1	3	2	1	1	1	3	1	1	1	1	3	2	2
3	I	2	2	1	3	1	2	1	3	3	2	3	1	1	2	3	3
	II	1	3	2	2	2	1	2	3	2	2	1	1	3	2	1	1
	III	3	1	3	1	1	1	2	1	1	1	2	3	1	1	2	3
4	I	3	2	2	1	2	3	2	2	1	1	1	3	3	3	2	2
	II	2	1	3	2	3	3	1	1	2	3	3	1	1	3	3	2
	III	1	1	1	3	2	3	3	3	1	2	3	1	3	1	2	3
5	I	3	1	2	2	1	1	2	2	3	1	2	2	1	3	1	2
	II	1	2	3	3	2	2	3	2	2	2	3	1	3	2	3	1
	III	2	3	1	1	1	1	3	3	2	3	2	3	2	1	2	3
6	I	2	3	2	1	3	2	1	3	3	2	3	2	2	1	1	3
	II	1	1	3	3	2	1	1	2	1	2	3	2	3	1	3	1
	III	3	2	1	2	1	1	3	1	3	3	1	1	2	2	3	2
7	I	1	2	2	3	3	1	1	3	1	3	1	2	1	2	3	2
	II	2	1	3	2	1	1	2	3	2	2	1	3	1	3	2	3
	III	3	3	1	1	3	1	3	3	1	2	3	1	2	1	3	3
8	I	3	2	1	2	1	2	2	1	2	3	1	1	3	3	1	1
	II	2	2	2	3	3	1	1	1	3	1	2	3	1	1	1	3
	III	3	3	3	1	1	3	2	2	3	1	1	2	2	2	2	1
9	I	3	1	2	2	3	1	3	2	2	1	2	2	2	3	1	2
	II	2	2	3	1	2	3	3	1	2	3	2	2	3	2	1	1
	III	1	3	1	3	3	3	3	2	1	3	1	3	3	1	2	3
10	I	3	1	1	3	2	1	3	1	3	3	1	1	3	3	1	1
	II	2	3	2	2	1	3	2	1	3	3	1	3	1	2	3	3
	III	1	2	3	1	3	1	2	2	1	1	2	2	2	2	2	2
11	I	3	1	2	3	1	2	1	2	3	3	3	2	3	3	2	1
	II	2	3	1	2	3	3	3	1	3	3	2	1	2	2	3	2
	III	1	2	3	1	2	1	2	3	2	2	1	3	3	3	2	3
12	I	2	2	3	1	2	2	1	3	2	1	3	3	1	2	2	3
	II	1	3	2	2	2	3	1	1	3	1	2	1	3	1	1	2
	III	3	1	1	3	1	2	3	2	2	3	1	2	2	3	3	1

Table A4 Designs with Four Alternatives

Choice Set	Profile	S: Standard Logit Attributes				M1: Mixed Logit Attributes				M2: Mixed Logit Attributes				M3: Mixed Logit Attributes			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	I	1	1	2	3	1	2	2	1	1	1	1	3	1	2	3	2
	II	1	1	3	2	1	3	3	1	1	2	2	1	3	1	2	1
	III	3	2	3	1	1	3	1	3	1	1	3	1	1	3	1	3
	IV	2	3	2	2	3	3	1	1	3	1	1	1	2	3	1	3
2	I	1	3	2	3	1	1	3	3	1	3	1	3	1	3	1	1
	II	1	2	3	1	2	1	2	3	1	1	3	1	3	1	1	2
	III	3	2	1	2	2	3	1	1	2	2	1	3	1	1	1	3
	IV	2	1	3	2	3	1	3	3	3	3	1	3	1	1	3	1
3	I	1	1	2	2	3	2	3	3	1	2	2	3	3	1	2	3
	II	2	2	2	2	3	1	3	2	1	2	3	2	2	1	3	1
	III	1	3	3	1	3	1	3	1	1	3	3	1	3	1	2	1
	IV	3	1	1	3	1	2	1	3	2	2	3	3	1	3	1	2
4	I	2	1	3	1	2	1	1	1	1	3	1	1	3	3	1	1
	II	3	1	2	1	3	1	1	1	2	1	1	2	1	2	2	3
	III	2	3	1	2	3	2	1	1	1	1	1	3	2	1	2	2
	IV	1	2	2	3	3	1	2	2	2	1	2	2	3	2	1	3
5	I	2	1	3	3	3	2	2	3	3	2	2	3	2	3	1	1
	II	3	2	2	2	1	2	2	2	2	3	1	2	1	1	3	3
	III	3	3	1	1	2	2	2	3	3	2	2	1	1	2	2	2
	IV	1	2	2	2	1	1	3	3	1	1	3	3	3	2	3	1
6	I	1	3	1	2	1	1	1	2	3	1	2	2	3	2	1	2
	II	2	2	3	1	1	2	1	1	2	3	2	1	2	1	3	2
	III	1	1	2	3	2	3	1	2	3	3	2	2	2	3	2	2
	IV	1	3	1	1	1	3	2	1	3	2	3	1	2	2	1	3
7	I	1	3	3	2	2	2	2	2	2	1	1	3	3	1	1	1
	II	2	2	1	3	1	2	1	2	2	2	2	2	1	3	2	1
	III	1	2	2	2	1	2	1	3	3	1	3	2	1	2	1	2
	IV	3	1	2	3	1	3	1	2	2	3	3	2	1	2	3	2
8	I	2	2	2	1	2	1	2	3	1	2	2	3	3	3	3	1
	II	2	2	1	3	2	2	1	2	2	3	3	1	2	2	3	1
	III	2	3	2	1	3	2	1	1	3	1	2	2	3	1	2	2
	IV	3	1	3	2	2	3	2	3	3	3	2	3	1	3	1	3
9	I	3	3	2	1	1	1	2	2	2	1	2	1	1	1	3	3
	II	2	1	1	3	1	2	3	2	1	2	3	2	2	2	2	3
	III	2	2	1	2	2	1	3	1	1	1	3	1	2	2	2	1
	IV	1	2	3	3	1	1	1	2	1	2	1	2	3	1	2	2

References

Allenby, Greg M., Neeraj Arora, James L. Ginter. 1998. On the heterogeneity of demand. *J. Marketing Res.* 35 (August) 384–389.

——, Peter J. Lenk. 1994. Modeling household purchase behavior with logistic normal regression. *J. Amer. Statist. Assoc.* 89 (428) 1218–1231.

——, ——. 1995. Reassessing brand loyalty, price sensitivity, and merchandising effects on consumer brand choice. *J. Bus. Econom. Statist.* 13 (3) 281–289.

——, Peter Rossi. 1999. Marketing models of consumer heterogeneity. *J. Econom.* 89 (1) 57–78.

Arora, Neeraj, Joel Huber. 2001. Improving parameter estimates and model prediction by aggregate customization in choice experiments. *J. Consumer Res.* 28 (September) 273–283.

Brownstone, David, Kenneth Train. 1999. Forecasting new product penetration with flexible substitution patterns. *J. Econom.* 89 (1–2) 109–129.

Bunch, David S., Jordan J. Louviere, Don Anderson. 1994. A comparison of experimental design strategies for multinomial logit models: The case of generic attributes. Working paper, Graduate School of Management, University of California, Davis, CA.

DeSarbo, Wayne S., Venkatram Ramaswamy, Steven H. Cohen. 1995. Market segmentation with choice-based conjoint analysis. *Marketing Lett.* 6 (2) 137–148.

Goldengorin, Boris, Gerard Sierksma. 1999. The data-correcting algorithm for the minimization of supermodular functions. *Management Sci.* 45 (11) 1539–1552.

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

Hamada, M., H. F. Martz, C. S. Reese, A. G. Wilson. 2001. Finding near-optimal Bayesian experimental designs via genetic algorithms. *Amer. Statist.* 55 (3) 175–181.

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

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

Kamakura, Wagner A., Michel Wedel, Jagadish Agrawal. 1994. Concomitant variable latent class models for conjoint analysis. *Internat. J. Res. Marketing* 11 (5) 451–464.

Kuhfeld, Warren F., Randall D. Tobias, Mark Garratt. 1994. Efficient experimental design with marketing research applications. *J. Marketing Res.* 31 (November) 545–557.

Lazari, Andreas G., Donald A. Anderson. 1994. Designs of discrete choice experiments for estimating both attribute and availability cross effects. *J. Marketing Res.* 31 (August) 375–383.

Lenk, Peter J., Wayne S. DeSarbo, Paul E. Green, Martin R. Young. 1996. Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Sci.* 15 (2) 173–191.

Louviere, Jordan J., George Woodworth. 1983. Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *J. Marketing Res.* 20 (November) 350–367.

McFadden, Daniel, Kenneth Train. 2000. Mixed MNL models of discrete response. *J. Appl. Econom.* 15 (Suppl.) 447–470.

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

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

Tang, Boxin. 1993. Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.* 88 (424) 1392–1397.

Wedel, Michel, Wagner A. Kamakura, Neeraj Arora, Albert Bemmaor, Jeongwen Chiang, Terry Elrod, Rich Johnson, Peter Lenk, Scott Neslin, Carsten Stig Poulsen. 1999. Discrete and continuous representation of heterogeneity. *Marketing Lett.* 10 (3) 217–230.

Zacks, S. 1977. Problems and approaches in design of experiments for estimation and testing in non-linear models. D. R. Krishnaiah, ed. *Multivariate Analysis, IV.* North-Holland, Amsterdam, 209–223.

This paper was received February 13, 2001, and was with the authors 7 months for 2 revisions; processed by Greg Allenby.