

Q ] ! [ çã \* Á@Á ~ãã } & Á ~ã áããã ~ãã ^áÁ^•ã } •Á | Á@  
{ ã^áÁ [ \* ã&@ ã^Á [ á^|Á^ Á & | áã \* Á [ çããã^•  
T ÉÔ:æà^Áã áÁ ÉXæ á^à | ^\

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

# Improving the efficiency of individualized designs for the mixed logit choice model by including covariates

M.Crabbe<sup>a,1,\*</sup>, M.Vandebroek<sup>a,b</sup>

<sup>a</sup>*Katholieke Universiteit Leuven, Faculty of Business and Economics, Naamsestraat 69, 3000 Leuven, Belgium.*

<sup>b</sup>*Leuven Statistics Research Center, W. de Croylaan 54, 3001 Leuven-Heverlee, Belgium.*

---

## Abstract

Recent research shows that the inclusion of choice related demo- and sociographics in discrete choice models aids in modeling the choice behavior of consumers substantially. However, the increase in efficiency gained by accounting for covariates in the design of a choice experiment has thus far only been demonstrated for the conditional logit model. Previous findings are extended by using covariates in the construction of individualized Bayesian  $\mathcal{D}$ -efficient designs for the mixed logit choice model. A simulation study illustrates how incorporating covariates affecting choice behavior yields more efficient designs and more accurate estimates and predictions at the individual level. Moreover, it is shown that the possible loss in design efficiency and therefore in estimation and prediction accuracy from including choice unrelated respondent characteristics is negligible.

*Keywords:* covariate, discrete choice experiment, mixed logit choice model, individual efficient design, hierarchical Bayes estimation

---

## 1. Introduction

Choice-based conjoint studies, also referred to as discrete choice or stated choice analyses, have become an established tool to gain insight in the choice behavior of consumers and are nowadays extensively applied in various research fields like marketing, transportation, environmental, health and political studies. For several years now, discrete choice research has focused on the value of covariates in describing and analyzing the preferences of consumers (Allenby et al., 2005; Boatwright et al., 2004; Bujosa et al., 2010; Colombo et al., 2009; Greene et al., 2006; Jen et al., 2003; Orme and Howell, 2009; Owen and Videras, 2009; Rose and Black, 2006; Rose and Bliemer, 2006). These researchers argue that, besides the attributes of the product profiles in the choice sets, covariates like demographics, socio-economic data or other individual-specific characteristics, influence respondents' choice behavior. Age, employment, income level, brand or store loyalty are just a few examples of variables that might affect choices. In order to describe preferences better, it is likely that taking these respondent characteristics into account in the setup and the analysis of the discrete choice experiment will be highly beneficial.

This research introduces the inclusion of covariates in the construction of individualized designs for the mixed logit or the random-effects discrete choice model. This model assumes preference heterogeneity and considers individual-specific parameters instead of an aggregate set. We use individual designs for the mixed logit choice model since these elicit choice data from each individual more efficiently in a heterogeneous population than an aggregate design (Yu et al., 2011). Unfortunately, the increase in design efficiency resulting from an individualized approach is accompanied by an increase in computational demand and logistic difficulties. It was the advent of online web surveys and impressive improvements in the processing speed of modern computers that made the construction of individual-specific designs for the mixed logit choice model technically feasible. Moreover, with web surveys it is very easy to obtain covariate information which can then be used for generating the individual designs. Since covariates might explain (part of) the heterogeneity in preferences, it is worthwhile to verify whether this inclusion improves design efficiency.

The advantages of including covariate information in discrete choice analysis have already been elaborately

---

\*Corresponding author

*Email addresses:* [marjolein.crabbe@econ.kuleuven.be](mailto:marjolein.crabbe@econ.kuleuven.be) (M.Crabbe), [martina.vandebroek@econ.kuleuven.be](mailto:martina.vandebroek@econ.kuleuven.be) (M.Vandebroek)

<sup>1</sup>Tel: +32 16 326961 / Fax: +32 16 326624

discussed in the literature for various discrete choice models. Allenby et al. (2005) and Orme and Howell (2009) for example estimate a mixed logit choice model incorporating covariate information in the mean of the random-effects distribution. Parameter estimates are obtained by means of a hierarchical Bayes estimation approach. Allenby et al. (2005) show how demographics and sociographics like age, gender and income influence preferences of credit card attributes and how they divide the population into covariate-based consumer segments with divergent choice behavior. Orme and Howell (2009) get a deeper insight in the preference structure of consumers buying a computer by asking each respondent about the amount they expect to spend during a next purchase and using this individual-specific information in the model to obtain more accurate purchase predictions. The incorporation of covariates in the random-effects distribution of mixed logit choice models was extended by Greene et al. (2006) and Rose and Black (2006), who not only include covariates in the means but also in the variances. Choice data concerning transport modes and mobile phone contracts are used respectively to illustrate the value of this model.

Covariates have also appeared to be valuable in latent class models, in which preference heterogeneity in a population is modeled by discrete latent classes instead of a random-effects distribution. Colombo et al. (2009) and Owen and Videras (2009) predict class membership in a latent class model by means of a multinomial logit formulation with demographics, like age and gender, and sociographics, like education level and income, determining the membership probabilities. Bujosa et al. (2010) consider a latent class choice model using covariates for class membership prediction together with random-effects in each segment. As such they combine a discrete and continuous approach to analyze preference heterogeneity.

Finally, Rose and Bliemer (2006) recommend the use of covariate information not only in the model specification but also in the experimental design for discrete choice models. They show an increase in design efficiency by including the covariates that are used in the analysis and estimation stage also in the design stage. They consider gender and income, categorized in three levels, in a study of transportation mode choice and incorporate these variables in the construction of efficient designs for the conditional logit model. In their research, a homogeneous and a heterogeneous design strategy taking the covariates into account are contrasted with a design approach that ignores the covariate information. In the homogeneous approach, they construct a single  $\mathcal{D}$ -efficient design for all respondents in the sample by incorporating covariate information in the  $\mathcal{D}$ -efficiency criterion. Applying a heterogeneous design strategy on the other hand, an efficient discrete choice experiment is designed for each of the six covariate segments separately. Their conclusions are twofold. First they demonstrate how both the homogeneous and heterogeneous design approach outperform the design strategy not taking the covariates into account, proving the benefit of incorporating covariates in the experimental design. Further, from their analysis it appears that more efficient choice data is obtained by designing experiments for each covariate segment separately as the heterogeneous design strategy results in larger absolute  $t$ -ratios for the model parameters than the homogeneous approach.

The findings of Rose and Bliemer (2006) are restricted to the conditional logit model. This research incorporates covariates not only in the heterogeneity distribution and the estimation of the mixed logit choice model but also in the construction of efficient individualized designs. The aim is to verify whether this improves the design efficiency and the accuracy of the estimates for the individual-specific parameters in the mixed logit choice model. A simulation study compares different combinations of design and analysis approaches with respect to their design efficiency and estimation and prediction accuracy. The different approaches are defined by whether or not individual designs are used to elicit choice data under the mixed logit choice model (in comparison to an aggregate design) and by whether or not covariates are included to set up the individualized designs and to analyze the corresponding choice data.

The remainder of this paper is organized as follows. The next section explains how one can incorporate covariates in the mixed logit choice model and its estimation. The use of covariate information to generate individually adapted efficient designs for the model is described in the third section. In the subsequent section, we first introduce the different design and analysis approaches. Next, a simulation study compares the approaches across various experimental settings and explores the covariate effects on efficiency and accuracy. Finally, section 5 contains some discussion and final conclusions.

## 2. Including covariates in the random-effects distribution of the mixed logit choice model

The mixed logit or the random-effects discrete choice model accounts for preference heterogeneity in a population as each person is assigned an individual-specific parameter vector  $\beta_n$ . Suppose every participant in a choice experiment evaluates  $S$  choice sets with  $K$  product profiles. McFadden's (1974) random utility model assumes that the utility that person  $n$  receives from alternative  $k$  in choice set  $s$  is the sum of a deterministic

observed part and a random unobserved part

$$U_{ksn} = \mathbf{x}'_{ksn} \boldsymbol{\beta}_n + \varepsilon_{ksn}, \quad (1)$$

with  $\mathbf{x}_{ksn}$  a  $p$ -dimensional vector capturing the attribute levels of the  $k$ th alternative shown to respondent  $n$  in choice set  $s$  and  $p$  the number of parameters in the model. The coefficients in  $\boldsymbol{\beta}_n$ , also referred to as partworths, express the main effects of the product features on utility for respondent  $n$ .

For each individual, the partworths are assumed constant over all choice sets in the experimental design. In this way, the mixed logit choice model accounts for correlation between an individual's successive choices and therefore mirrors the structure of a discrete choice experiment well. In essence, the mixed logit choice model describes each respondent's individual choice behavior with the conditional logit model. Consequently, conditional on  $\boldsymbol{\beta}_n$ , the probability that individual  $n$  chooses alternative  $k$  in choice set  $s$  equals (Train, 2003)

$$p_{ksn}(\boldsymbol{\beta}_n) = \frac{e^{\mathbf{x}'_{ksn} \boldsymbol{\beta}_n}}{\sum_{i=1}^K e^{\mathbf{x}'_{isn} \boldsymbol{\beta}_n}}. \quad (2)$$

The likelihood of respondent  $n$ 's series of choices  $\mathbf{y}_n^S$  for the  $S$  choice sets in the design is

$$L(\boldsymbol{\beta}_n | \mathbf{y}_n^S, \mathbf{X}_n^S) = \prod_{s=1}^S \prod_{k=1}^K (p_{ksn}(\boldsymbol{\beta}_n))^{y_{ksn}}, \quad (3)$$

with  $\mathbf{y}_n^S$  the choice information for respondent  $n$ , i.e.  $SK$  elements  $y_{ksn}$  which equal one if respondent  $n$  chooses profile  $k$  in choice set  $s$  and zero otherwise. The attribute levels  $\mathbf{x}_{ksn}$  of the  $SK$  product profiles in respondent  $n$ 's experimental design are stored in design matrix  $\mathbf{X}_n^S$ .

To model the choice behavior in the population, the mixed logit choice model assumes a random-effects distribution over the partworths  $\boldsymbol{\beta}_n$ . As mentioned in the introduction, not only the features of the product profiles but also specific covariates may influence consumers' choice behavior. Therefore, among others, Allenby et al. (2005) and Orme and Howell (2009) include covariate information in the mean of the random-effects distribution. We pursue their ideas in this research and assume the individual partworths in the mixed logit choice model depend on covariates according to the following regression model (Allenby et al., 2005; Greene et al., 2006; Orme and Howell, 2009; Rose and Black, 2006):

$$\boldsymbol{\beta}_n = \boldsymbol{\Theta} \mathbf{z}_n + \boldsymbol{\xi}_n, \quad (4)$$

with  $\boldsymbol{\xi}_n$  following a  $p$ -variate normal distribution  $\mathcal{N}(\boldsymbol{\xi}_n | \mathbf{0}, \boldsymbol{\Sigma})$ . The  $q$ -dimensional vector  $\mathbf{z}_n$  comprises the covariate values for respondent  $n$ , whereas the  $p \times q$  matrix  $\boldsymbol{\Theta}$  includes the regression parameters and expresses the sensitivity of the partworths with respect to the covariates. The individual-specific partworths thus follow a multivariate normal distribution  $\mathcal{N}(\boldsymbol{\beta}_n | \boldsymbol{\Theta} \mathbf{z}_n, \boldsymbol{\Sigma})$  with means  $\boldsymbol{\Theta} \mathbf{z}_n$  and covariance matrix  $\boldsymbol{\Sigma}$ . For categorical covariates, dividing the population into segments, different means are assumed for each subpopulation conditional on the levels of the covariates. The covariance matrix  $\boldsymbol{\Sigma}$  represents the heterogeneity around these segment means. The larger the diagonal elements of this covariance matrix, the more heterogeneity in consumer preferences remains unexplained.

Given the heterogeneity distribution  $\mathcal{N}(\boldsymbol{\beta}_n | \boldsymbol{\Theta} \mathbf{z}_n, \boldsymbol{\Sigma})$ , the unconditional likelihood of respondent  $n$ 's choices  $\mathbf{y}_n^S$  equals

$$\begin{aligned} L(\boldsymbol{\Theta}, \boldsymbol{\Sigma} | \mathbf{y}_n^S, \mathbf{X}_n^S, \mathbf{z}_n) &= \int L(\boldsymbol{\beta}_n | \mathbf{y}_n^S, \mathbf{X}_n^S) \phi(\boldsymbol{\beta}_n | \boldsymbol{\Theta} \mathbf{z}_n, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_n, \\ &= \int \left( \prod_{s=1}^S \prod_{k=1}^K (p_{ksn}(\boldsymbol{\beta}_n))^{y_{ksn}} \right) \phi(\boldsymbol{\beta}_n | \boldsymbol{\Theta} \mathbf{z}_n, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_n, \end{aligned} \quad (5)$$

with  $\phi$  the normal density. Note that in case covariates are not taken into account in the model specification, the random-effects distribution in (5) equals  $\phi(\boldsymbol{\beta}_n | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ , a multivariate normal density with mean vector  $\boldsymbol{\mu}$ .

To estimate both the hyperparameters  $\boldsymbol{\Theta}$  and  $\boldsymbol{\Sigma}$  and the individual coefficients  $\boldsymbol{\beta}_n$  in the mixed logit choice model, we apply an estimation approach combining Gibbs sampling with the Metropolis-Hastings algorithm (Allenby et al., 2005; Lenk et al., 1996; Orme and Howell, 2009; Yu et al., 2011). We follow Lenk et al. (1996) and Train (2003) and denote the approach as hierarchical Bayes estimation although very uninformative priors

are assumed. More details about the estimation of the mixed logit choice model including covariates are given in Appendix.

A great advantage of hierarchical Bayes estimation of random-effects models is that one is able to obtain good estimates for the respondent-specific partworths even when individual choice data is sparse (Allenby and Ginter, 1995; Allenby and Rossi, 1999; Jen et al., 2003; Lenk et al., 1996). The strength of hierarchical Bayes estimation is that it borrows data from the entire sample to obtain estimates at the individual level. It uses not only the individual choice data from the respondent but also draws information from the average choice behavior in the population through the random-effects distribution of the model parameters. Where other estimation techniques would only be able to produce aggregate results in case of insufficient data at the individual level, hierarchical Bayes estimation will rely more on the population level distribution (and less on the individual data) to obtain individual-level partworth estimates in this case. Several studies have illustrated the high estimation and prediction precision obtained from this approach (see for instance Allenby et al., 1995; Arora et al., 1998; Arora and Huber, 2001).

The hierarchical Bayes model shrinks the individual partworth estimates towards the population mean. In case covariates influence choice behavior and divide the population into covariate-based segments with segment-specific choice behavior, the assumption of a single normal population might create additional bias in the estimates. Incorporating the covariates in the mean of the heterogeneity distribution provides that, given the covariates, estimates are shrunk towards the mean  $\Theta \mathbf{z}_n$  of a more relevant subpopulation obtaining more accurate estimates at the individual level. The importance of parameter estimation at the individual level was stressed by Allenby and Ginter (1995). Product design, product modification and market segmentation are a few of many applications where the analysis of individual choice behavior proves valuable.

The use of covariates should not be restricted to the analysis of the choice data. It will appear fruitful to account for this additional information in the design of choice experiments too. The construction of individually adapted Bayesian efficient designs for the mixed logit choice model and the inclusion of covariates in this individualized design approach is discussed in the following section.

### 3. Incorporating covariates in the design construction to estimate the mixed logit choice model

Sándor and Wedel (2002) and Yu et al. (2009) construct locally and Bayesian  $\mathcal{D}$ -efficient designs respectively for a mixed logit choice model not taking the correlation between individuals' successive choices into account. Bliemer and Rose (2010) on the other hand do account for this correlation in the design setup but only manage to compute locally efficient designs. The intuitive subsequent step would be designing Bayesian efficient experiments for the mixed logit choice model incorporating the correlation between responses. Unfortunately, this turned out to be infeasible as it requires an extremely high computational effort.

In recent design literature, proper approximation of a model's information matrix is being discussed in order to optimize designs for complicated non-linear models, e.g. (first order) linearization, Gaussian quadrature and Laplace integral approximation. Such an approach sounds promising for our design problem but needs further research. We follow a different and innovative approach, introduced in Yu et al. (2011), and construct individual designs instead of an aggregate one to efficiently elicit choice data under the mixed logit choice model. Since preference heterogeneity is assumed, designs constructed at an individual level and adapted to the specific preference structure of a respondent are expected to be superior to designs constructed at an aggregate level. Aggregate designs are addressed to the average person in the sample and, as the preferences of the respondents may highly deviate from the average choice behavior, these designs may not be able to extract efficient choice data from all participants in the study.

The possibility to construct individually adapted designs leans on the structure of the mixed logit choice model in which the respondents' individual choice behavior is modeled according to the conditional logit model. Individual Bayesian  $\mathcal{D}$ -efficient designs are constructed to accurately estimate the underlying conditional logit models. These designs are set up in two stages, an initial static stage and an adaptive sequential stage (Danthurebandara et al., 2011; Yu et al., 2011). The initial stage comprehends the construction of a small initial individual design using common prior information about the model partworths. In the second stage choice sets are sequentially added to the initial design using Bayesian updates of the prior information based on the individual's observed choices. The procedure in Yu et al. (2011) is extended to include covariates in the prior mean. Using additional choice information, designs are fitted even more to the respondents' individual choice behavior, increasing their efficiency.

The individual designs are constructed with respect to the  $\mathcal{D}$ -criterion (for more information about this well known and widespread optimality criterion, see Atkinson et al., 2007). Minimization of the  $\mathcal{D}$ -error yields an efficient experimental design in the sense that this choice design minimizes the confidence region of the parameter

estimates obtained with choice data from this experiment. The  $\mathcal{D}$ -error equals the inverse of the determinant of the Fisher information matrix scaled by the power  $1/p$ , with  $p$  the number of parameters in the model. Note that we will use the generalized, instead of the ordinary Fisher information matrix of the conditional logit model as its inverse approximates the posterior covariance matrix better (Yu et al., 2008). The generalized Fisher information matrix considers the log-posterior density of the model parameters instead of the log-likelihood. For a given design  $\mathbf{X}$ , observed choices  $\mathbf{y}$  and prior distribution  $\pi(\boldsymbol{\beta}_n)$  for the model parameters, the matrix is given by:

$$\mathbf{I}_{GFIM}(\boldsymbol{\beta}_n, \mathbf{X}) = -\mathbf{E} \left[ \frac{\partial^2 \log[L(\boldsymbol{\beta}_n | \mathbf{y}, \mathbf{X}) \pi(\boldsymbol{\beta}_n)]}{\partial \boldsymbol{\beta}_n \partial \boldsymbol{\beta}_n'} \right]. \quad (6)$$

Assuming a multivariate normal prior with covariance matrix  $\boldsymbol{\Sigma}_0$  this gives

$$\begin{aligned} \mathbf{I}_{GFIM}(\boldsymbol{\beta}_n, \mathbf{X}) &= \mathbf{I}_{FIM}(\boldsymbol{\beta}_n, \mathbf{X}) + \boldsymbol{\Sigma}_0, \\ &= \sum_{s=1}^S \mathbf{X}_s' (\mathbf{P}_{sn} - \mathbf{p}_{sn} \mathbf{p}_{sn}') \mathbf{X}_s + \boldsymbol{\Sigma}_0, \end{aligned} \quad (7)$$

with  $\mathbf{I}_{FIM}(\boldsymbol{\beta}_n, \mathbf{X})$  the ordinary Fisher information matrix,  $\mathbf{X}_s$  the design matrix for choice set  $s$ ,  $\mathbf{P}_{sn} = \text{diag}(p_{1sn}, \dots, p_{Ksn})$  and  $\mathbf{p}_{sn} = (p_{1sn}, \dots, p_{Ksn})'$ .

Instead of fixing the coefficients  $\boldsymbol{\beta}_n$  in the criterion to specific prior values and obtaining locally efficient designs, uncertainty about the model parameters is taken into account in a Bayesian version of the  $\mathcal{D}$ -error. Bayesian  $\mathcal{D}$ -efficient designs are constructed by minimizing the expectation of the  $\mathcal{D}$ -error over the prior distribution  $\pi(\boldsymbol{\beta}_n)$ :

$$\mathcal{D}_B = \int \det[\mathbf{I}_{GFIM}(\boldsymbol{\beta}_n, \mathbf{X})]^{-1/p} \pi(\boldsymbol{\beta}_n) d\boldsymbol{\beta}_n. \quad (8)$$

In practice, the  $\mathcal{D}_B$ -error is approximated by the average  $\mathcal{D}$ -error over a sample of  $R$  draws  $\boldsymbol{\beta}_r$  from the prior distribution:

$$\tilde{\mathcal{D}}_B = \frac{1}{R} \sum_{r=1}^R \det[\mathbf{I}_{GFIM}(\boldsymbol{\beta}_r, \mathbf{X})]^{-1/p}. \quad (9)$$

The core of the individual design strategy is a recurring Bayesian update of the prior information  $\pi(\boldsymbol{\beta}_n)$  using a respondent's observed choice data. We assume the initial prior distribution  $\pi(\boldsymbol{\beta}_n)$  for the partworths  $\boldsymbol{\beta}_n$  is a multivariate normal  $\mathcal{N}(\boldsymbol{\beta}_n | \boldsymbol{\Theta}_0 \mathbf{z}_n, \boldsymbol{\Sigma}_0)$  with means  $\boldsymbol{\Theta}_0 \mathbf{z}_n$  and covariance matrix  $\boldsymbol{\Sigma}_0$ . Remark that we assume that covariate information is obtained before the actual choice experiment starts. The covariate values of the individual are stored in the vector  $\mathbf{z}_n$  and the matrices  $\boldsymbol{\Theta}_0$  and  $\boldsymbol{\Sigma}_0$  are prior values for the hyperparameters in the population obtained from a pilot study or previous experiments.

After designing an initial experiment  $\mathbf{X}_n^{S_1}$  with  $S_1$  choice sets for individual  $n$  and observing the corresponding choices  $\mathbf{y}_n^{S_1}$ , the prior information on respondent  $n$ 's partworths can be updated according to Bayesian rules, resulting in the following posterior distribution for  $\boldsymbol{\beta}_n$ :

$$q(\boldsymbol{\beta}_n | \mathbf{y}_n^{S_1}, \mathbf{X}_n^{S_1}, \mathbf{z}_n, \boldsymbol{\Theta}_0, \boldsymbol{\Sigma}_0) = \frac{L(\boldsymbol{\beta}_n | \mathbf{y}_n^{S_1}, \mathbf{X}_n^{S_1}) \phi(\boldsymbol{\beta}_n | \boldsymbol{\Theta}_0 \mathbf{z}_n, \boldsymbol{\Sigma}_0)}{\int L(\boldsymbol{\beta}_n | \mathbf{y}_n^{S_1}, \mathbf{X}_n^{S_1}) \phi(\boldsymbol{\beta}_n | \boldsymbol{\Theta}_0 \mathbf{z}_n, \boldsymbol{\Sigma}_0) d\boldsymbol{\beta}_n}, \quad (10)$$

with  $L(\boldsymbol{\beta}_n | \mathbf{y}_n^{S_1}, \mathbf{X}_n^{S_1})$  the likelihood of the choice data for the initial  $S_1$  choice sets and  $\phi$  the normal density.

The subsequent choice set  $\mathbf{x}_n^{S_1+1}$  for respondent  $n$  is found by considering all possible candidate sets and minimizing the  $\mathcal{D}_B$ -error in (8) of the combined design  $(\mathbf{X}_n^{S_1}, \mathbf{x}_n^{S_1+1})$  over the new prior  $q(\boldsymbol{\beta}_n | \mathbf{y}_n^{S_1}, \mathbf{X}_n^{S_1}, \mathbf{z}_n, \boldsymbol{\Theta}_0, \boldsymbol{\Sigma}_0)$ . Unfortunately, lacking a closed-form expression for this posterior distribution, there is no straightforward way to sample from it and therefore importance sampling is applied in the approximation of the  $\mathcal{D}_B$ -error (Albert, 2009; Bedrick et al., 1997; Yu et al., 2011). Moreover, instead of random draws, we use an 'intelligent' sample of extensible shifted lattice points for calculating the  $\tilde{\mathcal{D}}_B$ -error as this approximates the integral better (Yu et al., 2010; Yu et al., 2011).

Given the subsequent choice set  $\mathbf{x}_n^{S_1+1}$  and its corresponding response, the previously obtained prior infor-

mation  $q(\beta_n | \mathbf{y}_n^{S_1}, \mathbf{X}_n^{S_1}, \mathbf{z}_n, \Theta_0, \Sigma_0)$  can itself be updated

$$\begin{aligned} q(\beta_n | \mathbf{y}_n^{S_1+1}, \mathbf{X}_n^{S_1}, \mathbf{x}_n^{S_1+1}, \mathbf{z}_n, \Theta_0, \Sigma_0) &= \frac{L(\beta_n | y_n^{S_1+1}, \mathbf{x}_n^{S_1+1}) q(\beta_n | \mathbf{y}_n^{S_1}, \mathbf{X}_n^{S_1}, \mathbf{z}_n, \Theta_0, \Sigma_0)}{\int L(\beta_n | y_n^{S_1+1}, \mathbf{x}_n^{S_1+1}) q(\beta_n | \mathbf{y}_n^{S_1}, \mathbf{X}_n^{S_1}, \mathbf{z}_n, \Theta_0, \Sigma_0) d\beta_n}, \\ &= \frac{L(\beta_n | \mathbf{y}_n^{S_1+1}, \mathbf{X}_n^{S_1}, \mathbf{x}_n^{S_1+1}) \phi(\beta_n | \Theta_0 \mathbf{z}_n, \Sigma_0)}{\int L(\beta_n | \mathbf{y}_n^{S_1+1}, \mathbf{X}_n^{S_1}, \mathbf{x}_n^{S_1+1}) \phi(\beta_n | \Theta_0 \mathbf{z}_n, \Sigma_0) d\beta_n}, \end{aligned} \quad (11)$$

with  $y_n^{S_1+1}$  and  $\mathbf{y}_n^{S_1+1}$  respondent  $n$ 's choice data for the  $(S_1 + 1)$ th choice set and the first  $(S_1 + 1)$  choice sets respectively. Minimizing the  $\mathcal{D}_B$ -error over this new posterior enables us to find the  $(S_1 + 2)$ th choice set for respondent  $n$ . The recurring process of updating an individual's prior information by means of its observed choices and sequentially adding efficient choice sets is continued until a specific number of choice sets is reached.

The individual initial Bayesian  $\mathcal{D}$ -efficient design  $\mathbf{X}_n^{S_1}$  in (10) is obtained with the Bayesian modified Fedorov choice algorithm (Kessels et al., 2006; Zwerina et al., 1996). This algorithm begins with the construction of a candidate set of product profiles, in which the candidates equal all possible attribute level combinations. One gets a random starting design by randomly drawing profiles from this candidate set. Next, the random design is updated by exchanging the random design profiles with alternatives from the candidate set. An exchange is only carried out in case a decrease in Bayesian  $\mathcal{D}$ -error is observed. The exchange algorithm loops through all design profiles and is iterated multiple times to avoid ending up in local optima. The procedure is repeated for several starting designs, picking the most efficient updated design as the best in the end. As only covariate information is available at this initial stage,  $\mathcal{N}(\beta_n | \Theta_0 \mathbf{z}_n, \Sigma_0)$  is used as prior distribution in the construction of the initial efficient designs.

This two-stage design algorithm produces individually adapted Bayesian designs in a sequential way and uses choice data to update prior information about the model parameters. With these individual designs choice data is assembled to estimate the parameters in the mixed logit choice model. As covariates are included in the prior mean, the content of the initial prior is increased. We show that this results in individual designs eliciting choice data more efficiently than aggregate designs and individualized designs not using covariate information.

#### 4. Simulation study

This study aims to compare different design and analysis approaches in obtaining efficient designs and accurate individual-level parameter estimates and predictions for the mixed logit choice model and to verify whether the inclusion of covariates in design and/or analysis is valuable to this end. Further it is found out whether the incorporation of covariates can be harmful in case the covariates do not influence consumers' choice behavior. Therefore the effect on efficiency and precision including 'useless' covariates in design and analysis for the mixed logit choice model is studied as well. The simulation study consisted of two parts, differing in the number of covariates. In the first part the influence of only one binary covariate on consumers' choice behavior was assumed, whereas the second part dealt with two binary covariates.

We compare four different combinations of design and analysis approaches (Table 1). First, one can opt to use either individualized efficient designs or a traditional orthogonal design to obtain choice data under the mixed logit choice model. Further, the design and analysis approaches are defined by whether or not covariates are included in the construction of the individual designs and/or in the analysis of the choice data (and thus in the estimation of the mixed logit choice model).

	Design	Analysis
<i>C-C</i>	IASB designs with covariates	covariates
<i>NC-C</i>	IASB designs without covariates	covariates
<i>NC-NC(I)</i>	IASB designs without covariates	no covariates
<i>NC-NC(O)</i>	single nearly orthogonal design	no covariates

Table 1: Different design and analysis strategies

Consider first the design aspect of the different strategies in Table 1. The first three approaches generate individually adapted sequential Bayesian, IASB (Yu et al., 2011), designs for the mixed logit choice model. Yet, only the first approach incorporates covariates in the design construction. In this case covariates are included in the mean of the partworths' prior distribution. The second and the third approach also consider IASB designs,

but do not take the covariates into account to set up the choice sets. In contrast, in the fourth approach choice data is elicited from a single nearly orthogonal design which obviously is obtained without taking into account covariate information.

We choose all designs constructed in the simulations, i.e. both the individual-specific designs as the nearly orthogonal design, to comprise 16 choice sets with three alternatives. The product profiles in the sets are determined by three attributes, with three levels each. Effects type coding is used, so the attributes' levels correspond to (1,0), (0,1) and (-1,-1) respectively. Note that for this setup there does not exist an orthogonal design, that is why a nearly orthogonal design is used instead. For the individually adapted sequential Bayesian designs, five ( $S_1$ ) choice sets are computed in the initial static stage, the remaining 11 sets are sequentially added.

Not only in experimental design, but also in the analysis of the choice data one can choose whether or not to incorporate covariates. The first and the second design and analysis combinations include covariates in the estimation model, the third and fourth approach leave them out. Accounting for covariates in the design setup and the data analysis reflects the believe that they affect choice behavior and segment the market in consumer subpopulations. In contrast, excluding covariates from design and analysis assumes a single heterogeneous population. The design and analysis approaches in Table 1 are denoted by *C-C*, *NC-C*, *NC-NC(I)* and *NC-NC(O)* respectively. Note that *C* stands for *covariates*, *NC* for *no covariates* and *I* and *O* for the use of individualized efficient designs or a nearly orthogonal design respectively.

The design and analysis strategies will be compared across four different scenarios (I, II, III and IV), of which an overview is given in Table 2. The scenarios are determined by the setup of the experiment, more specifically by which respondents are used in the pilot study and by whether or not the true choice behavior of the consumers is influenced by the covariate(s).

	Study setup	Choice behavior
Scenario I	250 pilot (and main) respondents	influenced by covariate(s)
Scenario II		not influenced by covariate(s)
Scenario III	25 pilot respondents and	influenced by covariate(s)
Scenario IV	250 main respondents	not influenced by covariate(s)

Table 2: Different experimental setups considered in the simulation study

Each scenario consists of a pilot study and a main study. In the main study choice data is elicited from 250 respondents with either an individual design or an aggregate design and used to estimate their individual parameters in a mixed logit choice model (cfr. Table 1). Generating individualized designs requires prior values for the population parameters in the initial prior distribution, i.e.  $\Theta_0$  and  $\Sigma_0$  when covariates are incorporated in the design setup and  $\mu_0$  and  $\Sigma_0$  otherwise. We assume estimates for these hyperparameters are obtained by means of a pilot study (in which the respondents are presented the aggregate nearly orthogonal design) and consider two possible setups.

The first setup uses all the respondents from the main study also in the pilot study (study setup for scenario I and II in Table 2). In scenarios I and II there are thus 250 respondents in total, involved in both the pilot and the main study. They are denoted as pilot and main respondents in Table 2. Of course, it is not fully optimal to burden all participants with two choice experiments. They need to be called for or login to a server twice and need to go through a questionnaire twice, which increases the required effort from an individual. Unfortunately, when taking part in a survey takes too much of one's time, people tend to drop out. Therefore an additional setup is considered, using different individuals in the pilot and the main study. As most surveys deal with budget constraints, we consider only 25 respondents in the pilot study in this setup (study setup for scenario III and IV in Table 2). Scenarios III and IV thus assume 275 respondents in total, of which 25 participate in the pilot study and 250 in the main study. Note that the use of a nearly orthogonal design in the *NC-NC(O)* approach does not require prior information, therefore data obtained from the pilot study is also used to estimate the mixed logit choice model in this case.

For both experimental setups described above, two different cases will be considered (choice behavior in Table 2). The first case assumes that the covariate(s) influence(s) the choice behavior. The true  $\beta_n^*$ , used to simulate the choice data in the studies, is drawn from a multivariate normal distribution incorporating covariate information in the mean  $\mathcal{N}(\beta_n^*|\Theta^*z_n, \Sigma^*)$ . The parameters  $\Theta^*$  and  $\Sigma^*$  are the true (in reality unknown) population hyperparameters and are fixed to specific values in each simulation. In the second case it is assumed that choice behavior is not influenced by the covariate(s), so the true partworths are now simulated from a



multivariate normal with a single population mean  $\mathcal{N}(\beta_n^* | \mu^*, \Sigma^*)$ . Again  $\mu^*$  and  $\Sigma^*$  are preset to fixed values.

In all scenarios, the four design and analysis strategies, i.e. *C-C*, *NC-C*, *NC-NC(I)* and *NC-NC(O)*, will be used to obtain estimates  $\hat{\beta}_n$  for the individual-specific coefficients  $\beta_n^*$ . To evaluate and compare the results of the different approaches, some criteria are introduced in the following section to measure design efficiency and estimation and prediction accuracy.

#### 4.1. Design efficiency and estimation and prediction accuracy

As the individual designs are constructed with respect to the  $\mathcal{D}$ -optimality criterion, the  $\mathcal{D}$ -error is used to compare their efficiency. For each respondent, the local  $\mathcal{D}$ -error of his/her design  $\mathbf{X}_n$  is calculated with his/her specific  $\beta_n^*$ . Since efficiency of the designs is evaluated locally, the ordinary Fisher information matrix is used. Averaging over individuals gives a mean local  $\mathcal{D}$ -error

$$\mathcal{D}_L = \frac{1}{N} \sum_{n=1}^N \det[\mathbf{I}_{FIM}(\beta_n^*, \mathbf{X}_n)]^{-1/p}. \quad (12)$$

Obviously, for the design approach using an aggregate nearly orthogonal design,  $\mathbf{X}_n$  is the same for all individuals. Note that only the 250 individuals from the main study are considered to obtain values for the evaluation criteria.

Although the main focus in this research is on improving the efficiency of the individualized designs, also two related evaluation criteria measuring the estimation and prediction accuracy of the estimates at the individual level respectively are given. A root mean squared estimation error gives an indication of the average estimation error and is defined as follows

$$RMSE_{\beta} = \sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{\beta}_n - \beta_n^*)' (\hat{\beta}_n - \beta_n^*)}, \quad (13)$$

with  $\beta_n^*$  respondent  $n$ 's true individual-specific partworths and  $\hat{\beta}_n$  its estimate for a specific design and analysis approach. Note that for calculating the estimation errors, the full parameter vectors, including also the coefficients corresponding to the reference level of each attribute, are used. Obviously, in case of effects coding with three attribute levels, the coefficient corresponding to the reference (last) level is the negative sum of the two other coefficients.

Finally, analogous to the above, prediction accuracy of a design and analysis approach is evaluated by a root mean squared prediction error ( $RMSE_p$ ). To compute prediction errors, we consider the design including all possible choice sets with three alternatives. As in this research the product profiles in the experimental designs are determined by three three-level attributes (and therefore 27 possible profiles exist), this design comprises 2925, i.e.  $\binom{27}{3}$ , choice sets or in total 8775 ( $= 2925 \times 3$ ) profiles. For each individual the choice probabilities for all these product profiles are computed with the true and estimated partworths of the respondent and stacked in  $\mathbf{p}(\beta_n^*)$  and  $\mathbf{p}(\hat{\beta}_n)$  respectively. The root mean squared prediction error is obtained by comparing these vectors of choice probabilities and averaging over individuals:

$$RMSE_p = \sqrt{\frac{1}{N} \sum_{n=1}^N (\mathbf{p}(\hat{\beta}_n) - \mathbf{p}(\beta_n^*))' (\mathbf{p}(\hat{\beta}_n) - \mathbf{p}(\beta_n^*))}. \quad (14)$$

The remainder of this section discusses the simulation results (all simulations are implemented using the SAS procedure IML). Note that we only give an overview of the study including two covariates as similar results were obtained for the one-covariate case.

#### 4.2. The effect of two covariates on choice behavior

##### 4.2.1. Scenario I

In the first scenario we assume 250 individuals participate in the main study, after they completed a pilot study. Furthermore we assume consumers' choice behavior is truly affected by two covariates  $z_1$  and  $z_2$  and therefore simulate the individuals' true partworths from the heterogeneity distribution  $\mathcal{N}(\beta_n^* | \Theta^* \mathbf{z}_n, \Sigma^*)$ , with  $\mathbf{z}_n = [1, z_{1n}, z_{2n}]'$ . For now we choose the covariates to be dummy variables, taking the values -1 or 1. We

assume the true population hyperparameters  $\Theta^*$  and  $\Sigma^*$  are equal to

$$\Theta^* = \begin{pmatrix} 2 & 0.25 & 1.25 \\ 0 & 0 & 0 \\ 2 & 0.25 & 1.25 \\ 0 & 0 & 0 \\ 2 & 0.25 & 1.25 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \Sigma^* = \begin{pmatrix} 0.75 & -0.375 & 0 & 0 & 0 & 0 \\ -0.375 & 0.75 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.75 & -0.375 & 0 & 0 \\ 0 & 0 & -0.375 & 0.75 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.75 & -0.375 \\ 0 & 0 & 0 & 0 & -0.375 & 0.75 \end{pmatrix}.$$

The covariates divide the population into four consumer segments with distinct mean choice behavior. The partworth means of the covariate-based segments equal  $(0.5, 0, 0.5, 0, 0.5, 0)'$ ,  $(1, 0, 1, 0, 1, 0)'$ ,  $(3, 0, 3, 0, 3, 0)'$  and  $(3.5, 0, 3.5, 0, 3.5, 0)'$ . On average, all consumers in the population prefer the first level of the attributes to the second and the third level, but the strength of preference is very divergent between the segments. The covariate values for the respondents in the simulations are randomly chosen, resulting in approximately equally large segments. Note that for simulations with unequal segment sizes, similar results are obtained.

The variances in  $\Sigma^*$  are set to 0.75. Further we assume the correlation between coefficients corresponding to the same attribute is -0.5 (equivalently the covariance is -0.375). This ensures that the coefficient associated to the reference level of the attribute can be estimated with the same accuracy as the other levels' coefficients (Kessels et al., 2008; Vermeulen et al., 2010). The remaining correlations are zero. Note that the dimension of the individual-specific parameter vector  $\beta_n^*$  is 6, independent of the coding scheme used.

We ran 100000 Gibbs iterations to estimate each mixed logit choice model, considering 70000 draws for burn-in. For model estimation in the main study, we use the estimates from the pilot study as starting values for the population hyperparameters. For the individual partworths, starting values are set equal to the prior estimate of the mean, which is either  $\hat{\Theta}_0 \mathbf{z}_n$  in *C-C* and *NC-C* or  $\hat{\mu}_0$  in *NC-NC(I)* and *NC-NC(O)*. In contrast, when estimating a mixed logit choice model in a pilot study, there are no prior estimates available and starting values for the population mean parameters  $\Theta$  and  $\mu$  and all individual-level parameters are fixed to zero. The starting values for the population covariance matrix  $\Sigma$  are then set to  $p \times \mathbf{I}_p$  with  $p$  equal to 6, the number of parameters (Yu et al., 2011), corresponding to very uninformative starting information.

For each scenario, simulations were repeated five times and evaluation criteria were averaged over these five runs. Note that the simulations are very time consuming and therefore the number of runs was restricted to five. For scenario I, Table 3 shows the mean evaluation values with respect to design efficiency and estimation and prediction accuracy for the different design and analysis approaches. Standard deviations are given between brackets. Additionally, to visualize the differences between the approaches, the distributions of the individual local  $\mathcal{D}$ -errors and the individual root squared estimation and prediction errors (from a single run) are plotted in Figure 1.

	IASB with covariates ( <i>C-C</i> )	IASB without covariates ( <i>NC-C and NC-NC(I)</i> )	nearly orthogonal ( <i>NC-NC(O)</i> )
$\mathcal{D}_L$	0.274 (0.0032)	0.331 (0.0030)	0.575 (0.0174)

	<i>C-C</i>	<i>NC-C</i>	<i>NC-NC(I)</i>	<i>NC-NC(O)</i>
$RMSE_\beta$	1.430 (0.0326)	1.537 (0.0578)	2.008 (0.0560)	2.116 (0.0496)
$RMSE_p$	10.428 (0.2697)	10.999 (0.1090)	11.875 (0.0861)	13.417 (0.2675)

Table 3: Evaluation criteria (standard deviations) for the different design and analysis approaches (scenario I with two relevant covariates)

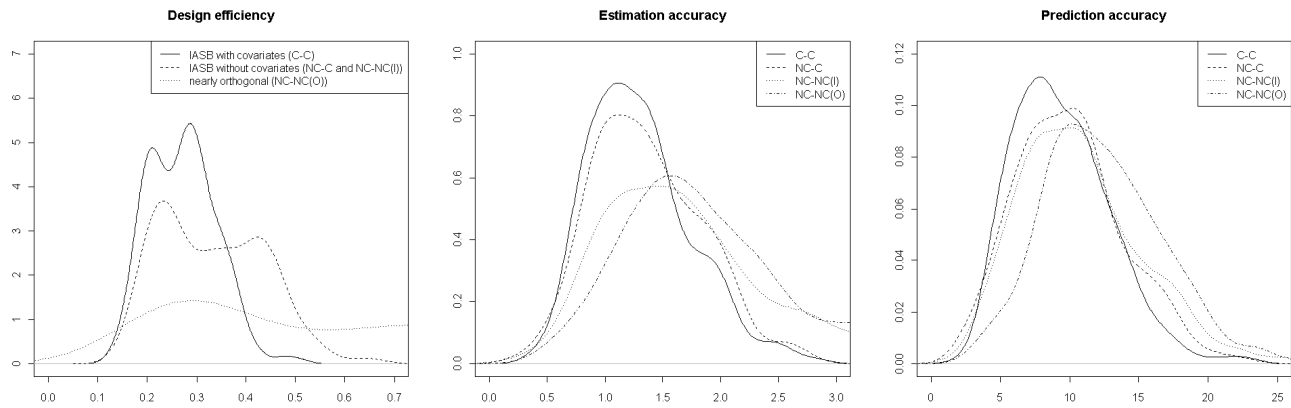


Figure 1: Distribution of the individual local  $\mathcal{D}$ -errors, root squared estimation and root squared prediction errors for the different design and analysis approaches (scenario I with two relevant covariates)

The inclusion of choice related covariates in the setup of individualized designs for the mixed logit choice model proves beneficial as the IASB designs generated with covariate information correspond to the significantly smallest mean local  $\mathcal{D}$ -error (0.274) and are therefore most efficient on average. Note that this design approach yields a decrease in average local  $\mathcal{D}$ -error of approximately 17% compared to IASB designs constructed without covariates. The  $C-C$  design and analysis approach also corresponds to the smallest mean estimation and prediction error, equal to 1.430 and 10.428 respectively and are significantly smaller than the corresponding values for  $NC-NC(I)$ . Clearly, when specific covariates influence choice behavior, one should include this information in the construction of individually adapted designs and in the specification and estimation of mixed logit choice model analyzing the choice data in order to obtain more efficient designs and more accurate estimates at the individual level. The superiority of including choice related covariates in both design and analysis is also reflected in the distributions of the individual errors, as this approach corresponds to the leftmost curves in all three plots. Finally, as already demonstrated in Yu et al. (2011), it is clear from the results that individualized efficient designs yield much more effective choice data than an aggregate orthogonal design.

Additionally, Figure 2 plots the distribution of the true simulated individual-specific partworths  $\beta_{1n}^*$ ,  $\beta_{2n}^*$  and  $\beta_{3n}^*$ , corresponding to the first attribute, together with the distribution of the partworth estimates for the  $C-C$ ,  $NC-C$ ,  $NC-NC(I)$  and  $NC-NC(O)$  design and analysis approaches for a single simulation. The covariates influence choice behavior, thus four consumer segments are present in the population. The vertical lines in the plots indicate the partworth means of the segments. Note that as some segment means lie rather close together, the distribution of the true partworths reveals only two segments. The distributions of the individual-level parameter estimates for the  $NC-NC(I)$  and the  $NC-NC(O)$  approach hardly resemble the true distribution. Moreover these design and analysis approaches fail to uncover the consumer segments in the population. The  $C-C$  and  $NC-C$  approaches detect the true distribution better, yet the  $C-C$  distribution of the estimates neighbors the true distribution best. This proves once more that one should include well considered covariates in design and analysis.

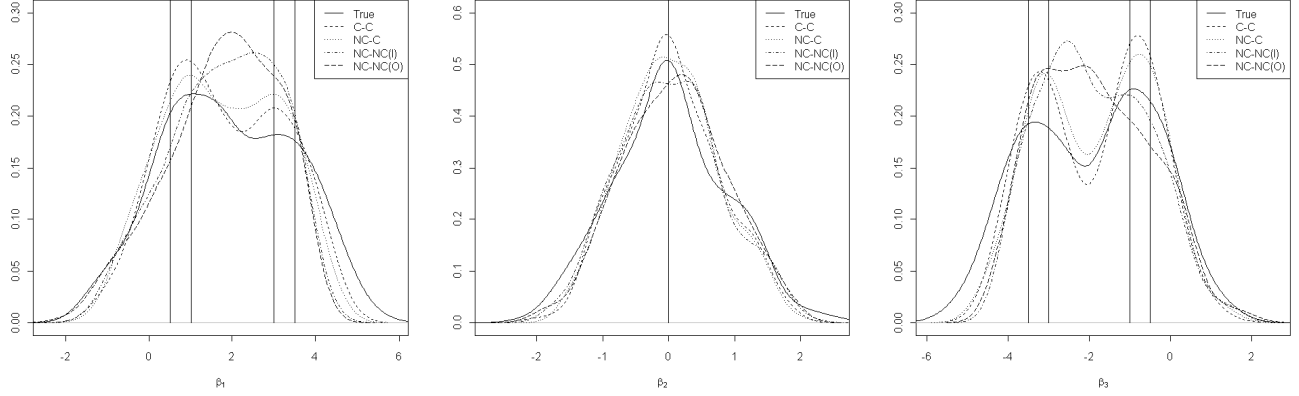


Figure 2: Distribution of the true individual-level partworths and the partworth estimates for the  $C-C$ ,  $NC-C$ ,  $NC-NC(I)$  and  $NC-NC(O)$  design and analysis approach

#### 4.2.2. Scenario II

In reality it is unknown whether covariates truly affect the consumers' preferences and whether it is beneficial to include them in the setup and analysis of a choice experiment. Therefore it is insufficient to restrict the study to covariates simulated to affect choice behavior. Scenario II, and later on scenario IV, assume and incorporate covariates that do not influence preferences. These scenarios verify whether, and if so how much, design efficiency and estimation and prediction accuracy deteriorate when covariates that do not explain the heterogeneity in preferences are included in design and analysis.

Similar to the preceding scenario, we consider the experimental setup with 250 participants. Yet, now it is assumed that choice behavior is not affected by covariates and that preference heterogeneity can be described by one normal distribution. The true individual partworths are simulated from  $\mathcal{N}(\beta_n^* | \mu^*, \Sigma^*)$  with  $\mu^*$  fixed to  $(2, 0, 2, 0, 2, 0)'$ . The variances in  $\Sigma^*$  are set to 1.75, the correlations between coefficients corresponding to the same attribute to -0.5. All remaining correlations equal zero. The averages of the evaluation criteria with respect to design efficiency and estimation and prediction accuracy are given in Table 4.

	IASB with covariates ( $C-C$ )	IASB without covariates ( $NC-C$ and $NC-NC(I)$ )	nearly orthogonal ( $NC-NC(O)$ )
$\mathcal{D}_L$	0.328 (0.0062)	0.377 (0.0036)	0.853 (0.0391)

	$C-C$	$NC-C$	$NC-NC(I)$	$NC-NC(O)$
$RMSE_\beta$	1.902 (0.0842)	1.977 (0.0564)	1.929 (0.0454)	2.240 (0.0392)
$RMSE_p$	10.624 (0.2510)	11.498 (0.0791)	11.548 (0.1197)	14.285 (0.1531)

Table 4: Evaluation criteria (standard deviations) for the different design and analysis approaches (scenario II with two superfluous covariates)

Even though the true choice behavior of the consumers is not affected by covariates in this scenario, it does not seem harmful to incorporate the 'useless' covariates in design and analysis. Individualized design including covariates still gives the smallest average local  $\mathcal{D}$ -error (0.328) and the  $C-C$  design and analysis approach still corresponds to the smallest mean estimation error (1.902), though the decrease in comparison to the  $NC-NC(I)$  approach is negligible, and the smallest mean prediction error (10.624). This remarkable result is probably due to the fact that with the extensive pilot study prior values for the population hyperparameters are estimated quite accurately. For  $\Theta_0$  this means that all elements related to the covariates are estimated to be almost zero and intercepts almost two, corresponding to no covariate effects. To see whether the use of covariates in design and analysis is also beneficial in case of little prior information, a smaller and therefore less informative pilot study is considered in scenario III and IV.

### 4.2.3. Scenario III and scenario IV

Both scenarios assume only 25 participants in the pilot study. Moreover these individuals are different from the 250 in the main study, resulting in less accurate and therefore less informative prior estimates for the hyperparameters in the heterogeneity distribution. Scenario III assumes preferences are affected by the covariates, in scenario IV the covariates are superfluous in modeling choice behavior. Table 5 and Table 6 summarize the simulation results for scenario III and scenario IV respectively. In addition, Figure 3 plots the distribution of the individual local  $\mathcal{D}$ -errors and the individual estimation and prediction errors from a single simulation in case the covariates influence choice behavior (scenario III).

	IASB with covariates ( <i>C-C</i> )	IASB without covariates ( <i>NC-C and NC-NC(I)</i> )	nearly orthogonal ( <i>NC-NC(O)</i> )
$\mathcal{D}_L$	0.303 (0.0078)	0.352 (0.0070)	0.608 (0.0226)

	<i>C-C</i>	<i>NC-C</i>	<i>NC-NC(I)</i>	<i>NC-NC(O)</i>
$RMSE_\beta$	1.511 (0.0596)	1.578 (0.0312)	2.073 (0.0952)	2.067 (0.0816)
$RMSE_p$	10.600 (0.2136)	11.305 (0.2731)	12.456 (0.4660)	13.048 (0.2682)

Table 5: Evaluation criteria (standard deviations) for the different design and analysis approaches (scenario III with two relevant covariates)

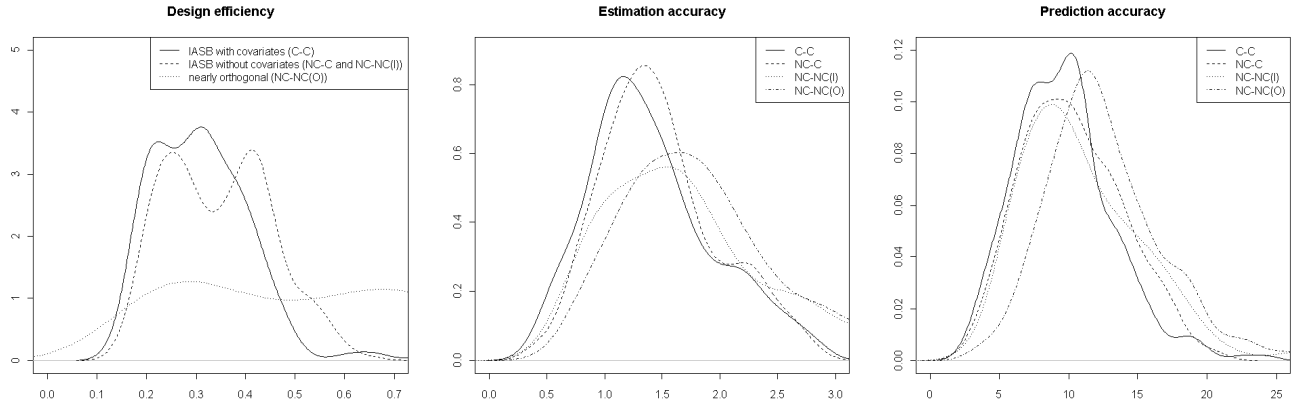


Figure 3: Distribution of the individual local  $\mathcal{D}$ -errors, root squared estimation and root squared prediction errors for the different design and analysis approaches (scenario III with two relevant covariates)

	IASB with covariates ( <i>C-C</i> )	IASB without covariates ( <i>NC-C and NC-NC(I)</i> )	nearly orthogonal ( <i>NC-NC(O)</i> )
$\mathcal{D}_L$	0.382 (0.0145)	0.393 (0.0101)	0.900 (0.0672)

	<i>C-C</i>	<i>NC-C</i>	<i>NC-NC(I)</i>	<i>NC-NC(O)</i>
$RMSE_\beta$	2.032 (0.0905)	1.999 (0.0567)	1.973 (0.0358)	2.292 (0.0277)
$RMSE_p$	11.522 (0.4797)	11.486 (0.1248)	11.558 (0.1274)	14.537 (0.2254)

Table 6: Evaluation criteria (standard deviations) for the different design and analysis approaches (scenario IV with two superfluous covariates)

Even with less accurate prior estimates from a much smaller pilot study, it still proves beneficial to incorporate choice related covariate information in the construction of individualized designs as the mean local  $\mathcal{D}$ -error of the IASB designs with covariates is significantly smaller than the one from the individual designs constructed without covariate information. Although the decrease in average local  $\mathcal{D}$ -error (approximately 14%) is smaller

than in scenario I, it is still noteworthy. Obviously, the smaller the pilot study, the lower the accuracy of the prior information and the lower the yield of incorporating covariates in the setup of individual efficient designs. Further, the *C-C* design and analysis approach still significantly outperforms the *NC-NC(I)* approach with respect to estimation and prediction accuracy.

In case choice behavior is not affected by the covariates (Tabel 6), logically, no significant increases in design efficiency or estimation and prediction accuracy are observed from including the covariates in experimental design and model specification and estimation. However, note that, just as in scenario II, it is not detrimental either to include a few useless covariates in design and analysis as potential decreases in efficiency and accuracy are negligible.

Until now, we have focused on categorical covariates dividing the population into segments with (possibly) divergent choice behavior. Inspired by the data experiment described in Allenby et al. (2005), we performed an additional simulation with both categorical and continuous covariates. Their data concerned a stated choice survey with credit card offers from regional banks. Besides the choices, customers provided information about their age (years), annual income (\$000) and gender (0 for male, 1 for female). In this additional study we considered a pilot study with 100 individuals and assumed the true choice behavior of the consumers is affected by the covariates. Besides one dummy variable, two more continuous covariates are influencing the choices. As we considered again product profiles defined by three three-level attributes, we only used the credit card attributes annual fee, bank and rebate. The true population hyperparameters making up the multivariate normal heterogeneity distribution of the individual-level partworths equal the estimates found in the paper. The findings are in line with the conclusions above. The most efficient designs and accurate estimates at the individual level are obtained by incorporating the covariates in the construction of the individualized designs and in the analysis of the choice data.

## 5. Discussion and conclusion

Recent research shows that demo- and sociographic information is not only useful for descriptive purposes in choice experiments but that, besides the individuals' choices, these variables might also hold information about one's choice behavior. This research includes covariates in the setup of efficient individual designs for the mixed logit choice model and in the analysis of the corresponding choice data increasing design efficiency and estimation and prediction accuracy.

The simulation results show that when specific covariates affect consumers' choice behavior, it is highly beneficial to incorporate these both in the individualized designs and in the specification and estimation of the mixed logit choice model. This approach corresponds on average to the highest design efficiency and to the smallest estimation and prediction errors. Obviously, at least some prior information about the effect of covariates on choice behavior in the population is necessary in order to attain benefits of incorporating choice related covariate information in the setup of individualized efficient designs. We show that the use of covariates with a pilot study of 25 individuals already improves the efficiency of the individualized designs substantially. Naturally, the larger the pilot study, the more efficient the designs but note that this also brings along additional costs.

This research confirms the value of covariates in discrete choice analysis and more specially, in individual experimental design and in hierarchical Bayes estimation of a mixed logit choice model including covariates in the heterogeneity distribution. Nevertheless, Orme and Howell (2009) advise not to add covariates rash and indiscriminately and alert for the presence of multicollinearity when including a lot of variables. One should only consider well thought covariates and only include covariates after their impact on choice behavior is verified. Note that for this decision also a pilot study can be used. Yet, in the worst case scenario, when covariates not affecting choice behavior are included in design and analysis, the potential loss in efficiency and accuracy appears to be insignificant. The use of choice unrelated covariates does not seem to be detrimental for design and estimation or, stated differently by Orme and Howell (2009), "there is evidence of general robustness to poor specification of covariates". However, this observation only holds for a limited number of superfluous covariates as we observed considerable decreases in efficiency and accuracy when using up to ten useless covariates in design and analysis. But if one is confident about the influence of a covariate on choice behavior, evidence is given that one should take it into account in individualized design for the mixed logit choice model and in analysis of the corresponding choice data in order to obtain more efficient designs and more accurate estimates and predictions at the individual level.

## Appendix. Hierarchical Bayes estimation of the mixed logit choice model with covariates

We follow Lenk et al. (1996) in choosing prior distributions for the hyperparameters  $\Theta$  and  $\Sigma$ . The prior for  $\Theta$  is assumed a  $p \times q$ -dimensional normal distribution  $N_{p \times q}(\mathbf{U}_0, \mathbf{V}_0)$  with mean  $\mathbf{U}_0$  and covariance matrix  $\mathbf{V}_0$ , which are  $p \times q$  and  $pq \times pq$  dimensional matrices respectively. Equivalently, the prior for  $\text{vec}(\Theta)$  equals  $N_{pq}(\text{vec}(\mathbf{U}_0), \mathbf{V}_0)$ , where  $\text{vec}(\Theta)$  and  $\text{vec}(\mathbf{U}_0)$  are obtained by stacking the columns of  $\Theta$  and  $\mathbf{U}_0$  respectively. We consider a non-informative prior for  $\Theta$  and therefore fix  $\mathbf{U}_0$  to  $\mathbf{0}_{p \times q}$  and  $\mathbf{V}_0$  to  $1000 \times \mathbf{I}_{pq}$ . The prior for  $\Sigma$  is a  $p$ -dimensional inverted Wishart distribution  $IW_p(\nu_0, \Delta_0^{-1})$  with  $\nu_0 = p$  degrees of freedom and scale matrix  $\Delta_0 = \mathbf{I}_p$ . To estimate the hierarchical model, a Markov chain Monte Carlo (MCMC) method is applied, more specifically Gibbs sampling combined with the Metropolis-Hastings algorithm. The Gibbs sampling algorithm makes use of the conditional posterior distributions for the model parameters and are given below.

Fix initial values  $\Theta^0$ ,  $\Sigma^0$  and  $\beta_n^0 \forall n$ . In the  $t$ th iteration of the Gibbs sampler, draws  $\Theta^t$ ,  $\Sigma^t$  and  $\beta_n^t \forall n$  are obtained using the following steps subsequently (Lenk et al., 1996; Train, 2003).

1. For each individual separately,  $\beta_n^t$  is obtained by means of the Metropolis-Hastings algorithm.
  - a. Consider the previous draw  $\beta_n^{t-1}$ .
  - b. Generate a  $p$ -dimensional vector  $\eta$  of independent standard normal values and compute the trial value  $\tilde{\beta}_n^t = \beta_n^{t-1} + \rho L \eta$ , with  $L$  the Choleski factor of  $\Sigma^{t-1}$  and  $\rho$  a scalar fixed by the researcher.
  - c. Compute the ratio

$$F = \frac{L(\tilde{\beta}_n^t | \mathbf{y}_n^S, \mathbf{X}_n^S) \phi(\tilde{\beta}_n^t | \Theta^{t-1} \mathbf{z}_n, \Sigma^{t-1})}{L(\beta_n^{t-1} | \mathbf{y}_n^S, \mathbf{X}_n^S) \phi(\beta_n^{t-1} | \Theta^{t-1} \mathbf{z}_n, \Sigma^{t-1})},$$

with  $L(\beta_n | \mathbf{y}_n^S, \mathbf{X}_n^S)$  the likelihood in (3) and  $\phi$  the normal density.

- d. Draw a value  $r$  from the standard uniform distribution. If  $r \leq F$  then  $\beta_n^t = \tilde{\beta}_n^t$ , else reject the trial value and  $\beta_n^t = \beta_n^{t-1}$ .
2. Draw  $\Theta^t$  from  $N_{p \times q}(\mathbf{U}, \mathbf{V})$  or draw  $\text{vec}(\Theta^t)$  from  $N_{pq}(\text{vec}(\mathbf{U}), \mathbf{V})$ , with

$$\begin{aligned} \mathbf{V} &= [\mathbf{Z}\mathbf{Z}' \otimes (\Sigma^{t-1})^{-1} + \mathbf{V}_0^{-1}]^{-1}, \\ \text{vec}(\mathbf{U}) &= \mathbf{V}[(\mathbf{Z} \otimes (\Sigma^{t-1})^{-1}) \text{vec}(\mathbf{B}) + \mathbf{V}_0^{-1} \text{vec}(\mathbf{U}_0)], \end{aligned}$$

and  $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_N)$ ,  $\mathbf{B} = (\beta_1^t, \dots, \beta_N^t)$ .

3. Generate  $\Sigma^t$  from  $IW_p(\nu, \Delta)$ , with

$$\nu = \nu_0 + N$$

and

$$\Delta = \frac{\Delta_0^{-1} + \sum_{n=1}^N (\beta_n^t - \Theta^t \mathbf{z}_n)(\beta_n^t - \Theta^t \mathbf{z}_n)'}{\nu}.$$

Iterating through these steps many times, the draws converge to the joint distribution of  $\Theta$ ,  $\Sigma$  and  $\beta_n \forall n$ . Means of the draws are used as estimates for the model parameters.

## Role of the funding source

Research funded by ZKC1090 / DBOF/08/014 - DBOF project of the KULeuven

## References

- Albert, J., 2009. *Bayesian Computation with R*. New York: Springer.
- Allenby, G.M., Arora, N., Ginter, J.L., 1995. Incorporating prior knowledge into the analysis of conjoint studies. *Journal of Marketing Research*. 32, 152-162.
- Allenby, G.M., Ginter, J.L., 1995. Using extremes to design products and segment markets. *Journal of Marketing Research*. 32, 392-403.
- Allenby, G.M., Rossi, P.E., 1999. Marketing models of consumer heterogeneity. *Journal of Econometrics*. 89, 57-78.
- Allenby, G.M., Rossi, P.E., McCulloch, R.E., 2005. Hierarchical Bayes models: a practitioners guide. Ssrn working paper, Ohio State University, University of Chicago.
- Arora, N., Allenby, G.M., Ginter, J.L., 1998. A hierarchical Bayes model of primary and secondary demand. *Marketing Science*. 17, 29-44.
- Arora, N., Huber, J., 2001. Improving parameter estimates and model prediction by aggregate customization in choice experiments. *Journal of Consumer Research*. 28, 273-283.
- Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. *Optimum experimental designs, with SAS*. Clarendon Press, Oxford U.K.
- Bedrick, E.J., Christensen, R., Johnson, W., 1997. Bayesian binomial regression: predicting survival at a trauma center. *The American Statistician*. 51, 211-218.
- Bliemer, M.C.J., Rose, J.M., 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B*. 44, 720-734.
- Boatwright, P., Dhar, S., Rossi, P.E., 2004. The role of retail competition, demographics and account retail strategy as drivers of promotional sensitivity. *Quantitative Marketing and Economics*. 2, 169-190.
- Bujosa, A., Riera, A., Hicks, R.L., 2010. Combining discrete and continuous representations of preference heterogeneity: a latent class approach. *Environmental and Resource Economics*. 47, 477-493.
- Colombo, S., Hanley, N., Louviere, J., 2009. Modeling preference heterogeneity in stated choice data: an analysis for public goods generated by agriculture. *Agricultural Economics*. 40, 307-322.
- Danthurebandara, V.M., Yu, J., Vandebroek, M., 2011. Sequential choice designs to estimate the heterogeneity distribution of willingness-to-pay. Forthcoming.
- Greene, W.H., Hensher, D.A., Rose, J., 2006. Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. *Transportation Research Part B*. 40, 75-92.
- Jen, L., Chou, C., Allenby, G.M., 2003. A Bayesian approach to modeling purchase frequency. *Marketing Letters*. 14, 5-20.
- Kessels, R., Goos, P., Vandebroek, M., 2006. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*. 43, 409-419.



- Kessels, R., Jones, B., Goos, P., Vandebroek, M., 2008. Recommendations on the use of Bayesian optimal designs for choice experiments. *Quality and Reliability Engineering International*. 24, 737-744.
- Lenk, P.J., DeSarbo, W.S., Green, P.E., Young, M.R., 1996. Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science*. 15, 173-191.
- McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, ed. Paul Zarembka. New York, Academic Press. 105-142.
- Orme, B., Howell, J., 2009. Application of covariates within Sawtooth Software's CBC/HB program: theory and practical example. Sawtooth Software conference papers. Sequim, WA: Sawtooth Software.
- Owen, A.L., Videras, J., 2009. Reconsidering social capital: a latent class approach. *Empirical Economics*. 37, 555-582.
- Rose, J., Black, I.R., 2006. Means matter, but variance matter too: decomposing response latency influences on variance heterogeneity in stated preference experiments. *Marketing Letters*. 17, 295-310.
- Rose, J., Bliemer, M.C.J., 2006. Designing efficient data for stated choice experiments. 11th International Conference on Travel Behavior Research, Kyoto.
- Sándor, Z., Wedel, M., 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Science*. 21, 455-475.
- Train, K., 2003. *Discrete choice methods with simulation*. Cambridge University Press.
- Vermeulen, B., Goos, P., Vandebroek, M., 2010. Obtaining more information from conjoint experiments by bestworst choices. *Computational Statistics and Data Analysis*. 54, 1426-1433.
- Yu, J., Goos, P., Vandebroek, M., 2008. A comparison of different Bayesian design criteria for constructing efficient conjoint choice experiments. Faculty of Business and Economics Research report. KBI-0817.
- Yu, J., Goos, P., Vandebroek, M., 2009. Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Science*. 28, 122-135.
- Yu, J., Goos, P., Vandebroek, M., 2010. Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments. *Transportation Research Part B*. 44, 1268-1289.
- Yu, J., Goos, P., Vandebroek, M., 2011. Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. *International Journal of Research in Marketing*. In press.
- Zwerina, K., Huber, J., Kuhfeld, W.F., 1996. A general method for constructing efficient choice designs, working paper, Fuqua School of Business, Duke University, Durham, NC 27708. Available from SAS Institute (<http://ftp.sas.com/techsup/download/technote/ts629.pdf>).