KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Fast algorithms to generate individualized designs for the mixed logit choice model

Marjolein Crabbe, Deniz Akınç, Martina Vandebroek

KBI_1303

Fast algorithms to generate individualized designs for the mixed logit choice model

Marjolein Crabbe [a1]

Email: marjolein.crabbe@kuleuven.be

Tel: +32 16 326961

Fax: +32 16 326624

Deniz Akınç [a]

Email: deniz.akinc@kuleuven.be

Martina Vandebroek [a,b]

Email: martina.vandebroek@kuleuven.be

[a] Faculty of Business and Economics, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium

[b] Leuven Statistics Research Centre, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium

[1] Corresponding author

Abstract

The mixed logit choice model has become the common standard to analyze transport behavior. Efficient design of the corresponding choice experiments is therefore indispensable to obtain precise knowledge of travelers' preferences. Accounting for the individual-specific coefficients in the model, this research advocates an individualized design approach. Individualized designs are sequentially generated for each person separately, using the answers from previous choice sets to select the next best set in a survey. In this way they are adapted to the specific preferences of an individual and therefore more efficient than an aggregate design approach. In order for individual sequential designs to be practicable, the speed of designing an additional choice set in an experiment is obviously a key issue. This paper introduces three design criteria used in optimal test design, based on Kullback-Leibler information, and compares them with the well-known $\mathcal{D}$-efficiency criterion to obtain individually adapted choice designs for the mixed logit choice model. Being equally efficient to $\mathcal{D}$-efficiency and at the same time much faster, the Kullback-Leibler criteria are well suited for the design of individualized choice experiments.

Keywords: discrete choice, mixed logit, individualized design, $\mathcal{D}$-efficiency, Kullback-Leibler information

1 Introduction

Discrete choice experiments are a popular and widely used survey methodology to study preferences in transportation (Axhausen et al., 2008; Bath, 2012; Hess and Hensher, 2010; Hess et al., 2008; Rose and Bliemer, 2009). Efficient design of these choice experiments is however essential to obtain precise estimates for the coefficients in the choice models. By efficiently selecting those choice sets that are most informative on the choice behavior, a higher level of estimation accuracy can be achieved for a given sample size, reducing the cost of the empirical study. In most researches, a single (or aggregate) design is used, which is equal for all respondents in the choice experiment. This study however continues on the recent developments on efficient individualized discrete choice design (Toubia et al., 2004; Yu et al., 2011).

In transportation research, the mixed (or random coefficients) logit choice model has been increasingly used for analyzing travelers' preferences (Bliemer and Rose, 2010; Greene et al., 2006; Hess and Hensher, 2010; Hess and Train, 2011). The strength of this model is that a heterogeneous population, and therefore individual-specific coefficients, are assumed. The model extends and improves the conditional logit choice model in which parameters are only estimated at an aggregate level. By accounting for preferences at the individual level, the mixed logit choice model mirrors real choice behavior better.

However, computing aggregate efficient designs for the mixed logit choice model is a lot more complicated. Bliemer and Rose (2010) were the first to construct aggregate $\mathcal{D}$-efficient designs for this model, but only succeeded in obtaining locally efficient designs assuming specific prior values for the coefficients in the model. Generating Bayesian aggregate $\mathcal{D}$-efficient designs, taking uncertainty about the model parameters into account, appeared infeasible in a reasonable amount of time. To circumvent the computational burden, Yu et al. (2011) introduced individualized Bayesian $\mathcal{D}$-efficient designs to elicit choice data for the mixed logit choice model. Note however that this alternative design approach is not only sensible because of technical boundaries. As the mixed logit choice model assumes individual-specific preferences and therefore individual-specific parameters, it is more in line with the underlying model assumptions to design individually adapted choice experiments instead of an aggregate design. Individualized choice designs are sequentially generated for each person separately by summarizing the answers to previous choice sets as prior information to efficiently select the next best set. By taking previous choices into account in the design process, the designs are tailored to the specific preferences of an individual. Therefore, individualized designs yield higher quality choice data and yield more efficient estimates for the mixed logit choice model than an aggregate design optimized for a simpler model (Danthurebandara et al., 2011; Yu et al., 2011).

As one can not let respondents wait for minutes, even seconds, obviously, sequential designs are only practicable if each additional set in the choice experiment is generated sufficiently fast. Despite the increasing computational capacity of modern computers, it thus remains necessary to search for algorithms that reduce the computation time of the design procedure. In this line, this research explores new design criteria that have been used in optimal test design to construct individually adapted choice designs for the mixed logit choice model and compares them with the $\mathcal{D}$-efficiency criterion that has often been used is this context.

In item response (or test) design, the individualized design approach has been generally accepted and successfully applied for years. It has become common practice to customize tests to the aptitude of a specific individual by incorporating a test taker's answers from previous test items to select the next best item in the test. Items too hard or too easy, adding hardly any information about an individual's ability, are in this way discarded from his/her test. Many of the test studies also apply $\mathcal{D}$-efficiency as optimality criterion, which is feasible in this context because of the simpler models involved. Recently however, three novel item selection rules, based on Kullback-Leibler information, have been introduced in the test design literature. The first maximizes the expected Kullback-Leibler divergence between subsequent posteriors of the individual-specific coefficients (Mulder and van der Linden, 2010; Wang and Chang, 2011). The other two criteria are derived from respectively mutual information (Mulder and van der Linden, 2010; Wang and Chang, 2011; Weissman, 2007) and entropy (Cheng, 2009; Wang and Chang, 2011), but are in essence also Kullback-Leibler distances.

For individualized test design, the new criteria have been shown to be very useful. Moreover, also in fields completely different to test theory these criteria appear to have great potential compared to traditional design approaches. Some examples are paired comparison designs for tournament scheduling (Glickman and Jensen, 2005), space-filling designs for computer experiments (Jourdan and Franco, 2010) and plasma diagnostics (Dreier et al., 2006). Encouraged by the positive results from the test design studies, we apply the Kullback-Leibler criteria to design individualized choice experiments. Their implementation in a discrete choice setting is shown to be efficient and very fast.

The remainder of this paper is organized as follows. The following section discusses the mixed logit choice model and the individualized design algorithms either employing the $\mathcal{D}$-error or the Kullback-Leibler information. Section 3 comprises an extensive simulation study comparing the efficiency and practicality of the design criteria. A final section closes the study with some conclusions.

2 Methodology

2.1 The mixed logit choice model

In a discrete choice experiment respondents must choose their preferred travel option in a series of choice sets contrasting multiple alternatives. Each alternative or profile in a set is characterized by a number of attributes, like for instance the travel time and the travel cost, taking specific values or levels. The mixed logit choice model probability that a person n chooses alternative k in choice set s (with K alternatives) equals

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

with $\mathbf{x}_{ksn}$ and $\boldsymbol{\beta}_n$ both p-dimensional vectors representing respectively the attribute levels of the kth alternative and individual n's coefficients. The latter expresses the individual's preferences with respect to, or alternatively stated the relative importance of, the attributes and their levels.

For a specific individual, the vector $\boldsymbol{\beta}_n$ is constant over all choice sets. The preferences of an individual are thus assumed not to vary across choice sets and are therefore essentially modeled by a conditional logit choice model. This insight was very important in the development of the individualized design approach for the mixed logit choice model in Yu et al. (2011) and applied later in this research.

Conditional on $\boldsymbol{\beta}_n$ and given the choice design $\mathbf{X}_n^S$ with S choice sets and corresponding choices $\mathbf{y}_n^S$, the likelihood of the model for respondent n is thus given by

$$L(\boldsymbol{\beta}_n | \mathbf{y}_n^S, \mathbf{X}_n^S) = \prod_{s=1}^{S} \prod_{k=1}^{K} [p_{ksn}(\boldsymbol{\beta}_n)]^{y_{ksn}}, \tag{2}$$

with $\mathbf{X}_n^S = (\mathbf{x}'_{11n}, ..., \mathbf{x}'_{K1n}, ..., \mathbf{x}'_{KSn})'$ the design matrix stacking the attribute levels of all profiles in the choice experiment and vector $\mathbf{y}_n^S$ comprising the elements y_{ksn} which are 1 if person n

chooses alternative k in choice set s and zero otherwise.

To model the aggregate choice behavior in the population the mixed logit choice model assumes a heterogeneity distribution, in most cases a multivariate normal distribution, over the individual-specific coefficients:

$$\boldsymbol{\beta}_n \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{3}$$

The unconditional likelihood for respondent n then equals

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{y}_n^S, \mathbf{X}_n^S) = \int L(\boldsymbol{\beta}_n|\mathbf{y}_n^S, \mathbf{X}_n^S)\ \phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma})\ d\boldsymbol{\beta}_n, \tag{4}$$

with ϕ the normal density.

In this research a Markov Chain Monte Carlo estimation approach, more specifically a Gibbs sampler, is used to estimate the mixed logit choice model (Lenk et al., 1996; Train, 2003; Yu et al., 2011).

2.2 Efficient individualized design for the mixed logit choice model

The following sections discuss the algorithms to construct individualized choice designs for the mixed logit choice model based on either $\mathcal{D}$-efficiency or Kullback-Leibler information. Note that although in essence the designs are optimized with respect to the underlying conditional logit choice models at the individual level, the choice data they yield is used to estimate a mixed logit choice model.

2.2.1 Minimum posterior weighted $\mathcal{D}$-error: a Fisher information design criterion

$\mathcal{D}$-efficient designs minimize the generalized variance of the parameter estimates (Atkinson et al., 2007), or equivalently, maximize the determinant of the model's Fisher information matrix, which is the negative expectation of the second partial derivative of the log-likelihood function. Assuming Bayesian estimation, the logarithm of the posterior will be used here instead of the logarithm of the likelihood yielding a Bayesian Fisher information matrix (BFIM). Given a design $\mathbf{X}_n^S$ for respondent n, this matrix equals:

$$\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^S) = -\mathrm{E}\left[\frac{\partial^2 \log[L(\boldsymbol{\beta}_n|\mathbf{y}_n^S, \mathbf{X}_n^S)\ f(\boldsymbol{\beta}_n)]}{\partial \boldsymbol{\beta}_n \partial \boldsymbol{\beta}_n'}\right], \tag{5}$$

with $L(\boldsymbol{\beta}_n|\mathbf{y}_n^S, \mathbf{X}_n^S)$ the likelihood in (2) and $f(\boldsymbol{\beta}_n)$ a prior distribution for individual n's coefficients. Note that the likelihood of the conditional logit choice model is applied as the choice designs are generated at the individual level. Assuming a multivariate normal prior with covariance matrix $\boldsymbol{\Sigma}_0$, the Bayesian Fisher information matrix becomes

$$\begin{aligned}
\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^S) &= \mathbf{I}_{FIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^S) + \boldsymbol{\Sigma}_0^{-1}, & (6)\\
&= \sum_{s=1}^{S} \mathbf{X}_{sn}'(\mathbf{P}_{sn} - \mathbf{p}_{sn}\mathbf{p}_{sn}')\mathbf{X}_{sn} + \boldsymbol{\Sigma}_0^{-1}, & (7)
\end{aligned}$$

with $\mathbf{I}_{FIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^S)$ the ordinary Fisher information matrix of the conditional logit choice model, $\mathbf{X}_{sn}$ the design matrix of choice set s, $\mathbf{P}_{sn} = \mathrm{diag}(p_{1sn}, ..., p_{Ksn})$ and $\mathbf{p}_{sn} = (p_{1sn}, ..., p_{Ksn})'$.

Instead of maximizing the determinant of this information matrix, we minimize the inverse, denoted as the $\mathcal{D}$-error and proportional to the volume of the confidence ellipsoid around the parameter estimates. Moreover, Bayesian $\mathcal{D}$-efficient (DB) designs are obtained, instead of locally efficient designs, by minimizing the expectation of the $\mathcal{D}$-error over a prior distribution of the individual-specific coefficients.

At the start of the choice experiment there is no choice data available. Therefore a multivariate normal prior, possibly the same one as in (5), is assumed and the following criterion is minimized over all possible choice sets to select the first set in the design

$$DB = \int \det[\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^1)]^{-1/p}\ f(\boldsymbol{\beta}_n)\ d\boldsymbol{\beta}_n, \tag{8}$$

with $f(\boldsymbol{\beta}_n) \equiv \phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$.

Yet, when a respondent has completed some choice sets, say $s-1$, the prior information can be updated in a Bayesian way with the choice data available. The posterior distribution of the individual-specific coefficients given the choices of the $s-1$ previous choice sets then equals

$$f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) = \frac{L(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, \mathbf{X}_n^{s-1})\ \phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)}{\int L(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, \mathbf{X}_n^{s-1})\ \phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)\ d\boldsymbol{\beta}_n}. \tag{9}$$

Note that $f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) \equiv f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, \mathbf{X}_n^{s-1}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, but the short form is applied for notational convenience.

This updated posterior is now used as the weighing distribution in the Bayesian $\mathcal{D}$-efficiency criterion to select the next best choice set for respondent n by minimizing

$$DB = \int \det[\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^s)]^{-1/p}\ f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\ d\boldsymbol{\beta}_n, \tag{10}$$

with $\mathbf{X}_n^s$ the design matrix including the $s-1$ perceived choice sets and the next sth choice set for respondent n. An additional set in an individualized choice experiment is thus obtained by minimizing the design's $\mathcal{D}$-error, weighted over the posterior distribution of the individual coefficients, hence "minimum posterior weighted $\mathcal{D}$-error". The process of alternately updating the posterior distribution of the coefficients with additional choice data and using this update to efficiently generate the next choice set can be continued until a specific amount of sets is administered. Finally note that instead of minimizing the posterior weighted $\mathcal{D}$-error, an alternative is to minimize the posterior weighted logarithm of the $\mathcal{D}$-error. This criterion was compared with (10) in Appendix A where it is shown that it yields equally efficient designs.

The following section introduces the alternative design criteria based on Kullback-Leibler divergence. In a simulation study the efficiency and practicality of the Kullback-Leibler criteria will be compared to that of the DB-criterion.

2.2.2 Kullback-Leibler information design criteria

The Kullback-Leibler divergence, also denoted as the Kullback-Leibler distance or the Kullback-Leibler information, between two density functions f and g of a continuous variable X is given by (Kullback and Leibler, 1951)

$$KL(f, g) = E_f\left[\log\frac{f(x)}{g(x)}\right] \tag{11}$$

$$= \int f(x)\,\log\frac{f(x)}{g(x)}\,dx. \tag{12}$$

It can be shown that for any f and g, KL is non-negative and zero in case of equal densities. Moreover, $KL(f, g)$ increases as the two densities become more divergent. That is why the Kullback-Leibler divergence is commonly interpreted as a measure of distance between two densities. Note however that KL is not a real distance measure as it is for instance not symmetric. This asymmetry will be important in the discrepancy between different design criteria in the remainder of this research.

Continuing the ideas of Chang and Ying (1996), who introduced Kullback-Leibler distance in optimal test design, Mulder and van der Linden (2010) developed an innovative selection rule to construct individualized designs by applying Kullback-Leibler divergence to the subsequent posteriors of an individual's model coefficients. More specifically, and applied to the discrete choice setting: in order to select the next best choice set for a specific respondent, one maximizes the distance between the current posterior of the coefficients (obtained with the choice data at hand) and the updated posterior one would obtain with the additional response information from the next choice set.

The criterion thus selects the choice set for which the response increases the information about the individual coefficients the most as the divergence between the subsequent posteriors is maximized. Since a set in a choice experiment comprises multiple alternatives (just as a test item might have multiple answers), we take the expectation over all possible choices and maximize the expected Kullback-Leibler distance between subsequent posteriors (Mulder and van der Linden, 2010).

Assume respondent n has completed $s - 1$ choice sets. The sth choice set in his/her design is then efficiently selected by maximizing

$$KLP = \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1})\ KL[f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}), f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})] \tag{13}$$

over all possible sets, with $f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})$ and $f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})$ updated posteriors as in (9). Note that y_{ksn} implies here that the kth alternative would be chosen in choice set s. The weights in (13) are defined as

$$\pi(y_{ksn}|\mathbf{y}_n^{s-1}) = \int p_{ksn}(\boldsymbol{\beta}_n)\ f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\ d\boldsymbol{\beta}_n \tag{14}$$

and equal the posterior weighted choice probabilities for the alternatives in the candidate set s, given previous choices. In Mulder and van der Linden (2010) this is denoted as the posterior predictive probability function.

To select the first choice set in the design, when no choice data is available yet, the current posterior $f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})$ in (13) and (14) is replaced by the normal prior $f(\boldsymbol{\beta}_n) \equiv \phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$:

$$\sum_{k=1}^{K} \pi(y_{k1n})\ KL[f(\boldsymbol{\beta}_n), f(\boldsymbol{\beta}_n|y_{k1n})] \tag{15}$$

with

$$\pi(y_{k1n}) = \int p_{k1n}(\boldsymbol{\beta}_n)\ \phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)\ d\boldsymbol{\beta}_n. \tag{16}$$

This is in analogy with (8). The same practice is used in all subsequent design algorithms.

Applying the definition of Kullback-Leibler distance from (12), one can show that KLP can be rewritten as

$$KLP = \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \left[\log \pi(y_{ksn}|\mathbf{y}_n^{s-1}) - \int \log p_{ksn}(\boldsymbol{\beta}_n)\ f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\ d\boldsymbol{\beta}_n\right]. \tag{17}$$

From the expression above it is clear that KLP only involves posterior weighted choice probabilities for the alternatives in the next set. This is in contrast to DB (10) where not only the next choice set but also all previous sets in the design are incorporated in the Fisher information matrix. Moreover, applying DB requires the computation of the determinant of this matrix, which makes the DB criterion a lot more complex than the KLP criterion. This highly influences the computation time of the algorithms, as illustrated later in this study.

Related to Kullback-Leibler divergence is mutual information, which for two continuous variables X and Y is defined as (Mulder and van der Linden, 2010; Weissman, 2007)

$$I_M(X,Y) = \int_Y \int_X f(x,y) \log \frac{f(x,y)}{f(x)f(y)} \, dxdy. \tag{18}$$

It is the Kullback-Leibler distance between the joint distribution of X and Y and their distribution in case of independence. It expresses how much information one variable holds with respect to the other. For instance, in case X and Y are independent, it is obvious that their mutual information is zero.

We follow Mulder and van der Linden (2010) and Wang and Chang (2011) and maximize the mutual information between the current posterior distribution of the individual coefficients and the posterior weighted choice probabilities for the alternatives in the next set, given the choice data of the previously administered sets. The choice set for which the response maximizes the information on the individual's coefficients is thus selected. The criterion to be maximized over all possible sets is

$$MUI = \sum_{k=1}^{K} \int f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1}) \log \frac{f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\pi(y_{ksn}|\mathbf{y}_n^{s-1})} \, d\boldsymbol{\beta}_n, \tag{19}$$

or equivalently,

$$MUI = \sum_{k=1}^{K} \left[\int p_{ksn}(\boldsymbol{\beta}_n) \log p_{ksn}(\boldsymbol{\beta}_n) \; f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) \; d\boldsymbol{\beta}_n - \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \log \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \right]. \tag{20}$$

From (20) it can be concluded that also MUI, just as KLP, only requires the computation of posterior weighted choice probabilities for the alternatives in the next choice set.

Moreover, the following calculations show how MUI and KLP are actually even more related as the MUI criterion in essence also maximizes expected Kullback-Leibler distance between posteriors. But, conversely to KLP for which the Kullback-Leibler distance between the current and the updated posterior is considered, the expected Kullback-Leibler distance between the updated and the current posterior is maximized now. As the Kullback-Leibler measure is not symmetric, MUI and KLP might be very similar but not equal. Yet, due to the high resemblance between both criteria we expect similar results with respect to their design efficiency.

$$\begin{aligned} MUI &= \sum_{k=1}^{K} \int f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1}) \log \frac{f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\pi(y_{ksn}|\mathbf{y}_n^{s-1})} \, d\boldsymbol{\beta}_n && (21) \\ &= \sum_{k=1}^{K} \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\pi(y_{ksn}|\mathbf{y}_n^{s-1}) \log \frac{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\pi(y_{ksn}|\mathbf{y}_n^{s-1})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\pi(y_{ksn}|\mathbf{y}_n^{s-1})} \, d\boldsymbol{\beta}_n && (22) \\ &= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log \frac{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})} \, d\boldsymbol{\beta}_n && (23) \\ &= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \; KL[f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}), f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})] && (24) \end{aligned}$$

The final design criterion used in this research is based on entropy. For a continuous variable X and density $f(x)$, the entropy is defined by (Wang and Chang, 2011; Weissman, 2007)

$$H(X) = -\int f(x) \log f(x) \, dx \qquad (25)$$

and is a measure of uncertainty. A trivial example is a Dirac delta distribution for which the entropy is zero. In contrast, entropy is maximal in case of a uniform distribution. To efficiently select a subsequent choice set in an individualized choice experiment, we minimize expected posterior entropy (Wang and Chang, 2011) or equivalently maximize

$$ENT = \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \, d\boldsymbol{\beta}_n. \qquad (26)$$

Similar to MUI and KLP, the entropy criterion can be written as the expected Kullback-Leibler distance between two different densities. To see this, note that (26) is equivalent to

$$ENT = \log c + \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log \frac{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})}{c} \, d\boldsymbol{\beta}_n, \qquad (27)$$

with c a constant. As log c can be discarded for maximization, maximizing ENT in (26) equals maximizing the second part in (27), which is the expected Kullback-Leibler distance between the updated posterior distribution of the coefficients and a uniform distribution. One can thus state that MUI has the current posterior $f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})$ as baseline in the Kullback-Leibler measure, whereas ENT has a uniform as baseline (Wang and Chang, 2011). This demonstrates how all three design criteria (KLP, MUI and ENT) are based on Kullback-Leibler distance and therefore highly related.

A final insight in the connection between the selection criteria, more specifically between MUI and ENT, follows from

$$
\begin{aligned}
MUI &= \sum_{k=1}^{K} \int f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1}) \log \frac{f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\pi(y_{ksn}|\mathbf{y}_n^{s-1})} \, d\boldsymbol{\beta}_n \qquad (28)\\
&= \sum_{k=1}^{K} \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\pi(y_{ksn}|\mathbf{y}_n^{s-1}) \log \frac{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\pi(y_{ksn}|\mathbf{y}_n^{s-1})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\pi(y_{ksn}|\mathbf{y}_n^{s-1})} \, d\boldsymbol{\beta}_n \qquad (29)\\
&= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \Bigg[\int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \, d\boldsymbol{\beta}_n - \\
&\qquad \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) \, d\boldsymbol{\beta}_n \Bigg] \qquad (30)\\
&= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \, [H(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) - H(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})]. \qquad (31)
\end{aligned}
$$

It is obvious that $H(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})$ is smaller than $H(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})$. Maximizing MUI thus selects the choice set that maximizes the expected decrease in uncertainty about an individual's coefficients. As the baseline in ENT is a uniform instead of the current posterior, it can be easily verified that $H(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})$ in (31) is replaced by a constant for the entropy criterion.

Due to the strong correspondence between KLP, MUI and ENT, it is expected that the differences in design efficiency for the mixed logit choice model between the methods will be small. For the comparison of the DB criterion with the Kullback-Leibler criteria on the other hand, it is less clear in advance whether or not one will outperform the other. Yet, regarding the computation time required to generate an additional choice set, we can expect the latter to be much faster than the former. DB involves the calculation of posterior weighted determinants of

Fisher information matrices which is far more complex than the measures required to compute KLP, MUI and ENT.

Note that the weighing of the design criteria over priors and posteriors is in practice approximated by weighted sums of these measures over draws from the distribution at hand. With a normal prior (at the beginning of the choice experiment), draws are easily obtained and given equal weight. Yet, sampling is more complex in the remainder of the experiment as there is no closed form for the updated posteriors of the coefficients. In this case importance sampling will be applied, more details are given in Appendix B.

3 Simulation study and results

It has already been elaborately demonstrated in Yu et al. (2011) and in Danthurebandara et al. (2011) that the use of individual sequential designs to estimate the mixed logit choice model is more efficient and yields more accurate estimates than an aggregate design optimized for a simpler model. Therefore, this study no longer focusses on proving the advantages of an individualized design approach but continues on these findings and considers alternative design criteria, beside $\mathcal{D}$-efficiency, to generate the individual designs. In this section, we compare the criteria DB, KLP, MUI and ENT introduced above with respect to their efficiency and practicality in designing individualized choice experiments for the mixed logit choice model. For generality, multiple experimental setups or scenarios are considered in the simulations, which differ regarding the number of attributes characterizing the travel options, the number of levels for each attribute and the number of alternatives in a choice set.

The first scenario assumes that all choice sets comprise two alternatives. Further, the profiles are characterized by three categorical attributes with three levels each. Designs with 15 choice sets are generated, the experimental setup can thus be displayed as $3^3/2/15$. Note that the algorithms are implemented such that a choice set can only be perceived once by a specific person. Further, due to long computation for the DB criterion, we assume only 50 respondents in the experiments. Some simulations were also performed with 200 respondents but as the main conclusions were the same we do not report these results. Moreover, we also want to find

	Experimental setup	$\boldsymbol{\mu}$	$\boldsymbol{\Sigma}$
Scenario 1	$3^3/2/15$	(-0.5, 0, -0.5, 0, -0.5, 0)	$0.5 \times \mathbf{I}_6$
Scenario 2	$2 \times 3 \times 2 \times 3/3/15$	(-0.5, -0.5, 0, -0.5, -0.5, 0)	$0.5 \times \mathbf{I}_6$
Scenario 3	$3 \times 2^4/2/15$	(2.403, 1.648, 0.976, -0.613, -0.188, 2.008)	Appendix C
Scenario 4	$3 \times 2 \times 3/3/15$	(0.419, 0.700, 1.355, 1.638, 1.690)	Appendix C

Table 1: Overview of the scenarios in the simulation study

out which criterion is most reliable when few choice data is available.

Effects coding is applied to the attribute levels, so the mean vector $\boldsymbol{\mu}$ and all $\boldsymbol{\beta}_n$ in the model include six coefficients. The true individual coefficients $\boldsymbol{\beta}_n$, used to simulate the choices in the simulations, are sampled from a normal heterogeneity distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, for which in this scenario $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are set respectively to (-0.5, 0, -0.5, 0, -0.5, 0) and $0.5 \times \mathbf{I}_6$, with $\mathbf{I}_6$ the 6-dimensional identity matrix. This implies that on average higher levels for the three attributes are preferred. An overview of the different scenarios considered in the simulation study is given in Table 1.

The choice sets in scenario 2 include three alternatives instead of two. Moreover, the profiles in these sets are now defined by four attributes with respectively 2, 3, 2 and 3 levels, again using effects coding. The hypothesized population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ equal (-0.5, -0.5, 0, -0.5, -0.5, 0) and $0.5 \times \mathbf{I}_6$. As also here designs with 15 choice sets are generated, this setup corresponds to $2 \times 3 \times 2 \times 3/3/15$.

Finally, the third and the fourth scenario are based on empirical studies from respectively Carlsson et al. (2003) and Espino et al. (2008). In scenario 3 the profiles are defined by five attributes, the first with three levels, the remaining four with two, and choice sets include two alternatives. In the final setup there are three attributes with respectively 3, 2 and 3 levels and three alternatives in each set. In both cases, choice experiments with 15 sets are designed. Note that in agreement with the assumptions in Carlsson et al. (2003) and in Espino et al. (2008) the attributes are now dummy coded. The values for $\boldsymbol{\mu}$ for both scenarios can be found in Table 1, the true values for $\boldsymbol{\Sigma}$ are given in Appendix C.

In all scenarios perfect prior information is assumed to construct the individually adapted choice designs. This means that the initial prior $f(\boldsymbol{\beta}_n)$ used in the design criteria is assumed to be a multivariate normal distribution for which the prior values $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ equal the true values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in each scenario. Obviously, less accurate estimates are obtained with miss-specified prior information, but it was found that the performance of the design criteria relative to one another remained the same, ensuring the robustness of the results. Finally note that to approximate the integrals in the selection rules, 512 draws are used for the results presented here.

3.1 Estimation and prediction accuracy

The main focus in the analysis of travelers' preferences is the aggregate choice behavior in a population, therefore we first discuss the accurateness of the estimates for the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in the mixed logit choice model. To compare the estimation accuracy obtained with the different design criteria, we compute the root-mean-squared-errors $\text{RMSE}_{\boldsymbol{\mu}}$ and $\text{RMSE}_{\boldsymbol{\Sigma}}$ which measure the estimation error for respectively $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. They are given by

$$\text{RMSE}_{\boldsymbol{\mu}} = \sqrt{\frac{(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})}{p}}, \tag{32}$$

with $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}$ respectively the estimates and the true values of the mean effects and p the number of coefficients in the model and

$$\text{RMSE}_{\boldsymbol{\Sigma}} = \sqrt{\frac{(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})'(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})}{p_{\boldsymbol{\Sigma}}}}, \tag{33}$$

with $\boldsymbol{\sigma}$ stacking all the unique elements from $\boldsymbol{\Sigma}$, $\hat{\boldsymbol{\sigma}}$ the estimates and $p_{\boldsymbol{\Sigma}}$ equal to $p(p+1)/2$, the number of elements in $\boldsymbol{\sigma}$.

Note that for each design algorithm and for each scenario, the generation of the choice designs and the estimation of the mixed logit choice model was repeated 100 times. The mean $\text{RMSE}_{\boldsymbol{\mu}}$ and $\text{RMSE}_{\boldsymbol{\Sigma}}$ values are given in Figure 1 for each scenario and all four design criteria. In the first two and the fourth scenario, no significant differences in $\text{RMSE}_{\boldsymbol{\mu}}$ and $\text{RMSE}_{\boldsymbol{\Sigma}}$ are observed. The population parameters in the model are estimated equally accurate with all four design criteria. In scenario 3 however, KLP outperforms the other methods as its $\text{RMSE}_{\boldsymbol{\mu}}$ and

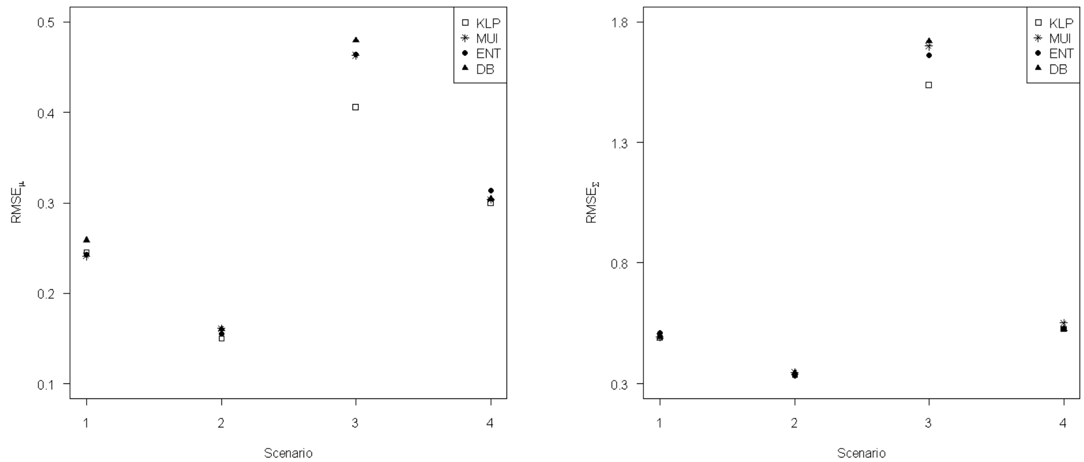

Figure 1: Mean $\text{RMSE}_{\boldsymbol{\mu}}$ and $\text{RMSE}_{\boldsymbol{\Sigma}}$ values obtained with KLP, MUI, ENT and DB for the different scenarios

$\text{RMSE}_{\boldsymbol{\Sigma}}$ are significantly smaller than the corresponding estimation errors obtained with MUI, ENT and DB.

The results above are a first indication that the Kullback-Leibler design criteria are worthy opponents of $\mathcal{D}$-efficiency as design criterion for the mixed logit choice model. The four design methods yield equally accurate estimates for the population parameters, though KLP occasionally shows smaller estimation errors. We are however cautious in generalizing this result as the increase in accuracy is quite small and not observed across all experimental setups.

Besides modeling aggregate choice behavior, it might also be of interest to have a view on individual preferences and to obtain good estimates for the coefficients $\boldsymbol{\beta}_n$ in the mixed logit choice model. Therefore, in addition to $\text{RMSE}_{\boldsymbol{\mu}}$ and $\text{RMSE}_{\boldsymbol{\Sigma}}$, the root-mean-squared-error $\text{RMSE}_{\boldsymbol{\beta}}$ is also compared between the design criteria:

$$\text{RMSE}_{\boldsymbol{\beta}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\frac{(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)}{p}}, \tag{34}$$

with $\hat{\boldsymbol{\beta}}_n$ and $\boldsymbol{\beta}_n$ respectively the estimates and the true values for the coefficients of person n and N the number of respondents. Figure 2 shows the mean values of the $\text{RMSE}_{\boldsymbol{\beta}}$ over the 100 simulations for each design criterion and each scenario.

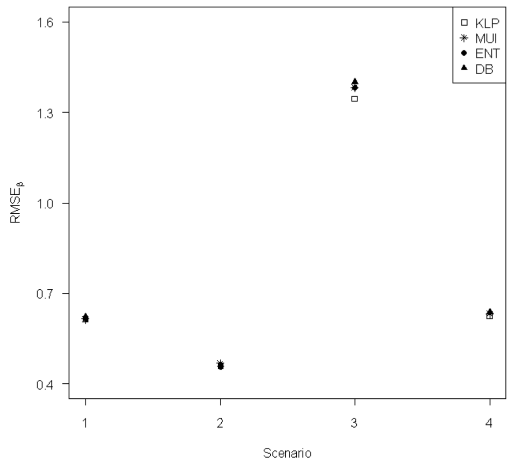

Figure 2: Mean RMSE$_{\boldsymbol{\beta}}$ values obtained with KLP, MUI, ENT and DB for the different scenarios

With respect to the individual coefficients, KLP again outperforms the remaining methods in scenario 3 as its RMSE$_{\boldsymbol{\beta}}$ value is significantly smaller than the estimation errors of MUI, ENT and DB. For all other scenarios, no clear differences in estimation error are observed. Figure 3 (left panel) displays density curves of the RMSE$_{\boldsymbol{\beta}}$ values in scenario 3 for each design criterion. The leftmost curve corresponds to KLP, confirming the on average smaller estimation errors. Such plots were also generated for the other scenarios, but as they only show overlapping density curves they are not given here.

Instead of averaging the estimation error over individuals as in (34), one can also compute the RMSE for each person separately,

$$\mathrm{RMSE}_n = \sqrt{\frac{(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)}{p}}. \tag{35}$$

In order to compare these individual estimation errors between methods, the RMSE$_n$ are plotted in ascending order (Figure 3, right panel). Note that these values again represent the averages of the individual errors over 100 simulations. We only show the plot for scenario 3, as the lines almost coincide for the other scenarios. The individual coefficients appear to be estimated equally well across the design criteria, yet for a large part of the respondents with

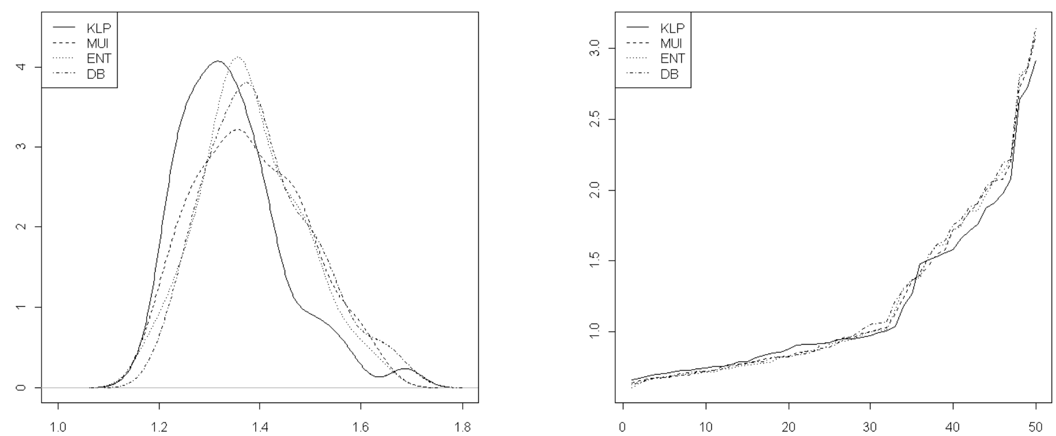

Figure 3: Density curves of the $\text{RMSE}_{\boldsymbol{\beta}}$ values (left) and mean RMSE_n values (right) obtained with KLP, MUI, ENT and DB for scenario 3

a high estimation error the RMSE_n values of the KLP criterion lie below the corresponding values for the other methods.

The quality of the estimates of the individual-specific coefficients can also be evaluated by measuring the accurateness of choice predictions. Therefore, in addition to the estimation errors above, a prediction error is composed. Consider vector $\mathbf{p}$, stacking the choice probabilities p_{ksn} of all profiles in all possible choice sets in a specific experimental setup. For scenario 1 for instance, the profiles are characterized by three three-leveled attributes. This gives 27 possible profiles and 351 possible choice sets with two alternatives. In the first scenario vector $\mathbf{p}$ thus includes 702 profiles. For scenario 2, 3 and 4 the number of possible choice sets is respectively 7140, 1128 and 816. As such, the number of profiles in $\mathbf{p}$ is respectively 21420, 2256 and 2448.

The prediction error is now defined as

$$\text{RMSE}_{\mathbf{p}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\frac{[\mathbf{p}(\hat{\boldsymbol{\beta}}_n) - \mathbf{p}(\boldsymbol{\beta}_n)]'[\mathbf{p}(\hat{\boldsymbol{\beta}}_n) - \mathbf{p}(\boldsymbol{\beta}_n)]}{S \times K}}, \tag{36}$$

with S the number of choice sets and K the number of alternatives in each set. As none of the scenarios reveal significant differences in prediction error between the design criteria, the values are not reported. Yet, these results confirm the equal efficiency of the different design criteria.

In conclusion, the simulations show that with choice data from individualized designs obtained with the Kullback-Leibler criteria both the population parameters and the individual-specific coefficients in the mixed logit choice model are estimated at least as accurate as with data from individualized Bayesian $\mathcal{D}$-efficient designs. Although optimizing different measures, the four design criteria thus perform equivalently. Note that this was expected for KLP, MUI and ENT as they are highly related. Yet, although overall evidence is lacking, slight preference could be given to KLP as design algorithm because of smaller estimation errors for some experimental setups.

3.2 Computation time

The main asset of the Kullback-Leibler design criteria is however that they are much easier to compute than the DB criterion. Although computing KLP, MUI and ENT also requires the weighing of choice probabilities over sequentially updated posteriors, the criteria do not involve the time consuming computation of the determinant of the Fisher information matrix, incorporating all sets in the choice experiment. Consequently, selecting the next best set in an individualized choice experiment is much faster with KLP, MUI and ENT than with DB.

To demonstrate this, Table 2 displays the average computation time (in seconds) to sequentially generate one additional set in a choice design using respectively KLP, MUI, ENT and DB as selection rule. Note that beside the 512 draws used in the simulations above to approximate the integrals in the criteria, calculations were also carried out with 1024 and 2048 draws.

The impressive decrease in computation time from using the Kullback-Leibler design criteria instead of DB is remarkable. Where for scenario 4 the DB computation times are approximately 18 times the KLP times, the computation times of the former even exceed more than 20 times the computation times of the latter for the other scenarios. Note, for instance, that in the second scenario using 512 draws, for which the selection of a choice set takes approximately only 1.8 seconds with KLP, MUI and ENT, the use of DB is far less attractive as each respondent must wait on average more than 35 seconds for every additional choice set.

	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
	512	1024	2048	512	1024	2048	512	1024	2048	512	1024	2048
KLP	0.074	0.152	0.285	1.726	3.381	6.729	0.207	0.402	0.805	0.219	0.418	0.809
MUI	0.082	0.152	0.293	1.773	3.484	6.924	0.215	0.414	0.816	0.223	0.426	0.836
ENT	0.090	0.168	0.328	1.972	3.866	7.674	0.242	0.468	0.914	0.246	0.473	0.934
DB	1.789	3.269	6.523	35.689	71.277	142.296	5.207	10.375	20.671	3.855	7.702	15.436

Table 2: Average computation time (seconds) for selecting one additional choice set with KLP, MUI, ENT and DB using various numbers of draws

However, the ratio of the run times between the methods is more important here than the absolute values as computing time can of course be reduced by using more powerful computers.

Amongst the novel criteria, KLP appears to be the fastest though the MUI criterion selects choice sets at about the same rate. ENT on the other hand, is slightly slower than KLP and MUI. This can be understood from the calculations below, showing that the computation of ENT requires computing an extra term over MUI. Although the differences in computation time are small, for more complex choice sets, with more alternatives and more and higher-leveled attributes, this may come into play.

$$
\begin{aligned}
ENT &= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\, d\boldsymbol{\beta}_n && (37)\\
&= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int \frac{p_{ksn}(\boldsymbol{\beta}_n)\, f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})}{\pi(y_{ksn}|\mathbf{y}_n^{s-1})} \log \frac{p_{ksn}(\boldsymbol{\beta}_n)\, f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})}{\pi(y_{ksn}|\mathbf{y}_n^{s-1})}\, d\boldsymbol{\beta}_n && (38)\\
&= \sum_{k=1}^{K} \int p_{ksn}(\boldsymbol{\beta}_n)\, f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) \log \frac{p_{ksn}(\boldsymbol{\beta}_n)\, f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})}{\pi(y_{ksn}|\mathbf{y}_n^{s-1})}\, d\boldsymbol{\beta}_n && (39)\\
&= \sum_{k=1}^{K} \int p_{ksn}(\boldsymbol{\beta}_n) \log p_{ksn}(\boldsymbol{\beta}_n)\, f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n - \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \log \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \\
&\quad + \sum_{k=1}^{K} \int p_{ksn}(\boldsymbol{\beta}_n) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n && (40)\\
&= MUI + \sum_{k=1}^{K} \int p_{ksn}(\boldsymbol{\beta}_n) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n && (41)
\end{aligned}
$$

Based on estimation and prediction accuracy no distinction could be made between the design criteria as they appeared equally efficient to estimate the mixed logit choice model. Comparing the complexity and consequently the speed of the criteria however, it is clear that the

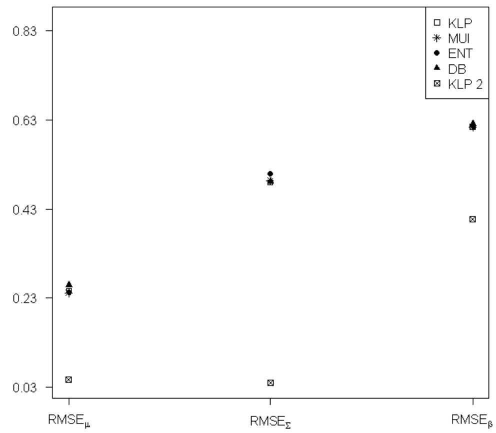

Figure 4: Mean RMSE$_{\boldsymbol{\mu}}$, RMSE$_{\boldsymbol{\Sigma}}$ and RMSE$_{\boldsymbol{\beta}}$ values obtained with KLP, MUI, ENT and DB for scenario 1 (cfr. Figure 1 and Figure 2) and obtained with KLP using designs with 20 choice sets for 1000 individuals and 2048 draws (KLP 2)

Kullback-Leibler design criteria, and more specifically KLP and MUI, are far more practicable than $\mathcal{D}$-efficiency.

An additional simulation was carried out for the first scenario, generating KLP choice designs with 20 sets for 1000 individuals. Moreover, 2048 draws were used instead of 512. Due to the speed of the KLP algorithm, the simulation could be run in a reasonable amount of time. Figure 4 gives the mean RMSE$_{\boldsymbol{\mu}}$, RMSE$_{\boldsymbol{\Sigma}}$ and RMSE$_{\boldsymbol{\beta}}$ values over 30 repetitions of this simulation, in addition to the estimation errors obtained before with the KLP, MUI, ENT and DB criteria (results for scenario 1 from Figure 1 and Figure 2). It can be concluded that, obviously, applying more data greatly improves the estimation accuracy.

3.3 Initial choice sets

To conclude this simulation section, we have to mention that in some cases efficient design for the mixed logit choice model could be slightly improved for DB by using initial sets in the choice experiments (Danthurebandara et al., 2011; Yu et al., 2011). The KLP, MUI and ENT algorithms were also implemented with initial sets, but for none of the scenarios considered they appeared beneficial. The use of initial sets implies that a small part, say S_I, of the total number of sets in the choice designs is generated in advance of collecting data in a non-sequential way using only the common prior information assumed for the individual coefficients. After the choices for these S_I initial sets are observed, the prior of the coefficients is updated a first time with the available choice data. The remainder of the sets in the choice experiment is then designed as described in section 2.2.1 by iteratively updating the posterior.

To select S_I initial sets $\mathbf{X}_n^{S_I}$ for respondent n with DB, we minimize

$$\int \det[\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^{S_I})]^{-1/p} \; f(\boldsymbol{\beta}_n) \; d\boldsymbol{\beta}_n, \tag{42}$$

with $f(\boldsymbol{\beta}_n) \equiv \phi(\boldsymbol{\beta}_n | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ the normal prior for the individual coefficients. Note that instead of selecting only one choice set at the beginning of the experiment as in (8), S_I choice sets are now incorporated in the initial design $\mathbf{X}_n^{S_I}$ and the corresponding Fisher information matrix $\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^{S_I})$.

The implementation of initial sets might be beneficial as the preference information inherent in choice data is small. This is not that much of a problem in case a number of sets has already been completed as the prior can then be updated with a sufficient amount of choice information, tailoring the choice design in an efficient way. But it might be problematic in the beginning of a choice experiment. Updating the prior for the first time after some predetermined initial sets might then be more beneficial than updating it after each set from the start.

For all four scenarios, DB designs (15 sets in total) were also generated with five initial sets. But only for scenario 3 the initial sets appeared beneficial, clearly reducing the estimation errors for $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}_n$. Note however that only the decreases in $\text{RMSE}_{\boldsymbol{\mu}}$ and $\text{RMSE}_{\boldsymbol{\beta}}$ (and not in $\text{RMSE}_{\boldsymbol{\Sigma}}$) were significant in comparison to the errors obtained with DB designs without initial sets. Moreover, even with the use of initial sets, DB is still not outperforming KLP, MUI or ENT.

4 Conclusion

In this study, we discussed the efficient design of choice experiments to obtain choice data for the mixed logit choice model. Despite the model's increasing popularity to analyze the preferences of travelers, the search for efficient designs for this model is still in its infancy. As the construction of aggregate designs for the mixed logit choice model appeared only feasible for local optimality criteria (Bliemer and Rose, 2010), Yu et al. (2011) introduced an individualized design approach applying $\mathcal{D}$-efficiency to estimate the model. Individualized choice experiments are designed with respect to the individual preferences of a specific respondent, sequentially taking previous choices into account to select the next choice set. With individual-specific coefficients in the model, designing choice experiments at an individual level is more efficient than using an aggregate design approach.

This research elaborates on these findings and focuses on improving the practicability of individualized choice design. Three new design criteria, alternative to $\mathcal{D}$-efficiency, are presented to speed up the construction of the individualized designs. The first criterion studied maximizes the expected Kullback-Leibler divergence between subsequent posteriors of the individual-specific coefficients, the second maximizes the mutual information between the current posterior of the individual coefficients and the choice probabilities for the alternatives in the next choice set and the third minimizes expected posterior entropy. Though defined by different concepts, all three criteria can be written as Kullback-Leibler information measures.

In a simulation study the Kullback-Leibler criteria are compared with $\mathcal{D}$-efficiency under various experimental settings. The scenarios differ with respect to the number of attributes characterizing the profiles, the number of levels for each attribute and the coding of these levels and the number of alternatives in each choice set. The conclusions however are unanimous: the design efficiency of the four criteria to estimate the mixed logit choice model is equivalent. Both the population parameters and the individual-specific coefficients in the model are estimated equally accurate with choice data from the four optimality algorithms. Yet, although not generally observed, the criterion maximizing expected Kullback-Leibler divergence between subsequent posteriors could be given slight preference as smaller estimation errors were some-times obtained.

The main result from this study is however that the Kullback-Leibler criteria are to be preferred to the traditional $\mathcal{D}$-efficiency criterion due to their low complexity, yielding a huge decrease in computation time. Selecting an additional choice set for an individual takes, for the experimental setups studied, approximately 20 times longer with the $\mathcal{D}$-criterion than with the Kullback-Leibler criteria. The alternative design criteria thus give researchers the opportunity to obtain more choice data, to consider more complex choice experiments or to improve the approximation of the criteria at a same (or even lower) computational cost. Where individualized design was the solution for efficiently designing choice experiments for the mixed logit choice model, the Kullback-Leibler criteria warrant the feasibility.

A The $DB_{\log}$ criterion

As stated in section 2.2.1, DB designs minimize the expectation of the inverse of the determinant of $\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^s)$. An alternative approach is to maximize the expected logarithm of the determinant of this information matrix or thus, minimize the posterior weighted logarithm of the $\mathcal{D}$-error:

$$DB_{\log} = \int \log\left(\det[\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^s)]^{-1/p}\right) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\; d\boldsymbol{\beta}_n.$$

The logarithmic transformation makes the DB optimality criterion less sensitive to very small and very large determinant values (Atkinson et al., 2007).

To compare the estimation accuracy of both criteria, the density curves of the $\text{RMSE}_{\boldsymbol{\beta}}$ values for scenario 1 are shown in Figure 5, the plots are similar for the remaining scenarios. The graphs illustrate that the DB and the $DB_{\log}$ criterion are equivalent in design efficiency. Similar plots have been obtained for $\text{RMSE}_{\boldsymbol{\mu}}$ and $\text{RMSE}_{\boldsymbol{\Sigma}}$.

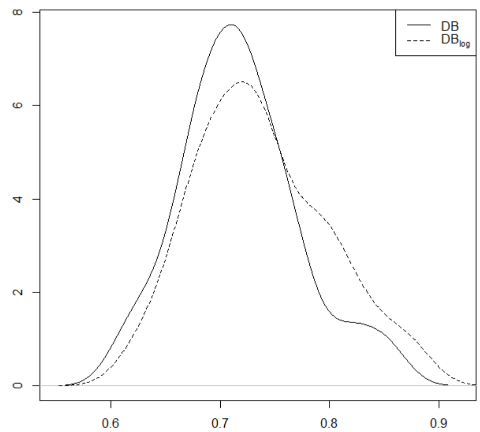

Figure 5: Density curves of the $\text{RMSE}_{\boldsymbol{\beta}}$ values obtained with DB and $DB_{\log}$ for scenario 1

B Importance sampling

Assume a measure C depending on β, weighted over a distribution $f(\beta)$ of β from which sampling is infeasible:

$$\int C(\beta)\ f(\beta)\ d\beta.$$

To approximate this integral, importance sampling applies an importance density $g(\beta)$ from which draws can be easily obtained. With R draws β^r from $g(\beta)$ the expression above is approximated by

$$\sum_{r=1}^{R} C(\beta^r)\ w_r,$$

with the importance weights w_r defined as

$$w_r = \frac{f^*(\beta^r)/g^*(\beta^r)}{\sum_{t=1}^{R} f^*(\beta^t)/g^*(\beta^t)}$$

and f^* and g^* the kernels of respectively f and g.

To approximate the design criteria in this research, the importance density for the posterior is assumed a multivariate student t distribution with the posterior mode as mean and variance-covariance matrix $-\mathbf{H}_M^{-1}$, with $\mathbf{H}_M$ the Hessian of the posterior $f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})$ evaluated at the mode (Yu et al., 2011). Note that instead of random sampling, extensible shifted lattice points transformed with Baker's transformation (Yu et al., 2010) are applied to obtain the draws from the importance density.

C Matrices Σ in scenario 3 and 4

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 6.605 & 6.784 & 3.299 & 2.059 & 2.246 & 1.855 \\ 6.784 & 8.231 & 4.838 & 3.018 & 3.290 & 2.721 \\ 3.299 & 4.838 & 7.007 & -0.547 & 2.424 & -0.186 \\ 2.059 & 3.018 & -0.547 & 5.392 & 2.339 & 2.344 \\ 2.246 & 3.290 & 2.424 & 2.339 & 3.964 & 1.663 \\ 1.855 & 2.721 & -0.186 & 2.344 & 1.663 & 9.358 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_4 = \begin{pmatrix} 0.047 & 0 & 0 & 0 & 0 \\ 0 & 0.906 & 0 & 0 & 0 \\ 0 & 0 & 2.632 & 0 & 0 \\ 0 & 0 & 0 & 0.568 & 0 \\ 0 & 0 & 0 & 0 & 1.107 \end{pmatrix}$$

Acknowledgements

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

References

Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. Optimum experimental designs, with SAS. Clarendon Press, Oxford U.K.

Axhausen, K.W., Hess, S., König, A., Abay, G., Bates, J.J., Bierlaire, M., 2008. Income and distance elasticities of values of travel time savings: New Swiss results. Transport Policy. 15, 173-185.

Bath, C.R., 2012. Recent developments in discrete choice model formulation, estimation and inference. Transportation Research Part B. 46, 273-275.

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.

Carlsson, F., Frykblom, P., Liljenstolpe, C., 2003. Valuing wetland attributes: an application of choice experiments. Ecological Economics. 47, 95-103.

Chang, H., Ying, Z., 1996. A global information approach to computerized adaptive testing. Applied Psychological Measurement. 20, 213-229.

Cheng, Y., 2009. When cognitive diagnosis meets computerized adaptive testing: CD-CAT. Psychometrika. 74, 619-632.

Danthurebandara, V., Yu, J., Vandebroek, M., 2011. Sequential choice designs to estimate the heterogeneity distribution of willingness-to-pay. Quantitative Marketing and Economics. 9, 429-448.

Dreier, H., Fischer, R., Dinklage, A., Hirsch, M., Kornejew, P., 2006. Optimization of plasma diagnostics using Bayesian probability theory. Bayesian Inference and Maximum Entropy Methods in Science and Engineering, 26th International Workshop.

Espino, R., Martín, J.C., Román, C., 2008. Analyzing the effect of preference heterogeneity on willingness to pay for improving service quality in an airline choice context. Transportation Research Part E. 44, 593-606.

Glickman, M.E., Jensen, S.T., 2005. Adaptive paired comparison design. Journal of Statistical Planning and Inference. 127, 279-293.

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, 7592.

Hess, S., Hensher, D.A., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. Transportation Research Part B. 44, 781-790.

Hess, S., Smith, C., Falzarano, S., Stubits, J., 2008. Managed-lanes stated preference survey in Atlanta, Georgia: measuring effects of different experimental designs and survey. Transportation Research Record. 2049, 144-152.

Hess, S., Train, K.E., 2011. Recovery of inter- and intra-personal heterogeneity using mixed logit models. Transportation Research Part B. 45, 973-990.

Jourdan, A., Franco, J., 2010. Optimal Latin hypercube designs for the Kullback-Leibler criterion. Advances in Statistical Analysis. 94, 341-351.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. The Annals of Mathematical Statistics. 22, 79-86.

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.

Mulder, J., van der Linden, W.J., 2010. Multidimensional adaptive testing with Kullback-Leibler information item selection. In: van der Linden, W.J., Glas, C.A.W. (eds.) Elements of adaptive testing. pp. 77-101. Springer, New York.

Rose, J.M., Bliemer, M.C.J., 2009. Constructing efficient stated choice experimental designs. Transport Reviews. 29, 587-617.

Toubia, O., Hauser, J.R., Simester, D.I., 2004. Polyhedral methods for adaptive choice-based conjoint analysis. Journal of Marketing Research. 41, 116-131.

Train, K., 2003. Discrete choice methods with simulation. Cambridge University Press.

Wang, C., Chang, H., 2011. Item selection in multidimensional computerized adaptive testing - gaining information from different angles. Psychometrika. 76, 363-384.

Weissman, A., 2007. Mutual information item selection in adaptive classification testing. Educational and Psychological Measurement. 67, 41-58.

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. 28, 378-388.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be