Transportation Research Part B 60 (2014) 1–15

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

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

Marjolein Crabbe [a,*], Deniz Akinc [a], Martina Vandebroek [a,b]

[a] Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium
[b] Leuven Statistics Research Center, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 29 October 2012
Received in revised form 16 November 2013
Accepted 21 November 2013

Keywords:
Discrete choice
Mixed logit
Individualized design
$\mathcal{D}$-efficiency
Kullback–Leibler information

ABSTRACT

The mixed logit choice model has become the common standard to analyze transport behavior. Moreover, more and more transport studies start to make use of stated preference data to obtain precise knowledge on travelers' preferences. Accounting for the individual-specific coefficients in the mixed logit choice model, this research advocates an individualized design approach to generate these stated choice experiments. 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. 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.

1. Introduction

Discrete choice is a popular and widely used methodology to study preferences in transportation (Axhausen et al., 2008; Bhat, 2012; Hess and Hensher, 2010; Hess et al., 2008; Rose and Bliemer, 2009). Although revealed preference is still most applied, more and more transport studies start to make use of stated preference data to analyze the transport behavior. In revealed preference adequate information concerning the attribute levels of the real-life choice options faced might be missing, which complicates the analysis of the choice behavior but which can be avoided by shifting to the use of stated preference choice experiments (Hess et al., 2007). It is then however beneficial to efficiently design these experiments to obtain precise estimates for the coefficients in the choice models. By 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 in efficient individualized discrete choice design (Bliemer and Rose, 2010b (for the conditional logit choice model); Toubia et al., 2004; Yu et al., 2011 (for the mixed logit choice model)).

In transportation research, the mixed (or random coefficients) logit choice model has been used since the nineties (Ben-Akiva et al., 1993; Bhat, 1998; Revelt and Train, 1998) and it is still popular to analyze travelers' preferences (Bliemer and Rose, 2010a; Greene et al., 2006; Hess and Hensher, 2010; Hess and Train, 2011). This model extends and improves the conditional logit choice model in which parameters are only estimated at an aggregate level. The strength of the mixed logit

* Corresponding author. Tel.: +32 16 326961; fax: +32 16 326624.
E-mail address: marjolein.crabbe@kuleuven.be (M. Crabbe).

0191-2615/$ - see front matter
http://dx.doi.org/10.1016/j.trb.2013.11.008

choice model is that, in addition to explaining preference heterogeneity with covariates for instance, the model is able to take unexplained heterogeneity into account. The mixed logit choice model assumes individual-specific coefficients following a heterogeneity distribution in the population and as such mirrors real choice behavior better. Moreover, as the individual coefficients are assumed constant throughout the choice experiment, the model accounts for the correlation between the respondents' successive choices.

However, computing aggregate efficient designs for the mixed logit choice model is a lot more complicated. Sándor and Wedel (2002) and Yu et al. (2009) obtained respectively locally and Bayesian aggregate $\mathcal{D}$-efficient designs for the cross-sectional mixed logit choice model which neglects the panel structure of the data. Bliemer and Rose (2010a) were the first to construct aggregate $\mathcal{D}$-efficient designs for the model considering the correlations between the individuals' choices, but only succeeded in obtaining locally efficient designs assuming specific prior values for the coefficients. Generating Bayesian aggregate designs for the panel version of the mixed logit choice model, 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 cannot 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 methods 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.

Kullback–Leibler divergence was originally introduced in adaptive test design by Chang and Ying (1996) in a search for more global design criteria, as at that point Fisher information was only applied as a local criterion by constructing efficient designs at intermediate estimates of the model coefficients. In the beginning of a test, with few data available, this could however be problematic. The interest in Kullback–Leibler divergence as design criterion continued to grow due to its easy generalization to a multidimensional setting and its straightforward interpretation as a distance measure (Mulder and van der Linden, 2010). The main goal of using Kullback–Leibler design criteria remained however the same: generating designs in an efficient way to estimate model coefficients as accurate as possible with a given amount of data.

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. In the mixed logit choice model, the probability that a person n chooses alternative k in choice set s (with K alternatives) equals

$$p_{ksn}(\boldsymbol{\beta}_n) = \frac{e^{\mathbf{x}'_{ksn}\boldsymbol{\beta}_n}}{\sum_{t=1}^{K} e^{\mathbf{x}'_{tsn}\boldsymbol{\beta}_n}}, \tag{1}$$

with $\mathbf{x}_{ksn}$ and $\boldsymbol{\beta}_n$ both $(p \times 1)$-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 (Revelt and Train, 1998; Yu et al., 2011). 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 cfr. Train (2003)

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

with $\mathbf{X}_n^S = (\mathbf{x}'_{11n}, \ldots, \mathbf{x}'_{K1n}, \ldots, \mathbf{x}'_{KSn})'$ the $(KS \times p)$-dimensional 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 over the individual-specific coefficients. In this research, a multivariate normal heterogeneity distribution is assumed:

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

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.

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). Note however that the model can straightforwardly be adapted to the use of alternative heterogeneity distributions such as lognormals, triangulars or constrained distributions, which are frequently applied in transportation studies to model the random coefficients (Bhat and Sidharthan, 2012; Hensher and Greene, 2003; Hess et al., 2005; Hess and Polak, 2005).

2.2. Efficient individualized design for the mixed logit choice model

The following sections discuss in detail 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. A summarizing overview of the algorithms is given in Appendix A (the corresponding SAS code is available at http://www.econ.kuleuven.be/martina.vandebroek/trb-code).

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. Yet, assuming Bayesian estimation for the underlying conditional logit choice models, the logarithm of the posterior is 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 (Berger, 1985):

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

$$= \mathbf{I}_{FIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^S) - E\left[\frac{\partial^2 \log[f(\boldsymbol{\beta}_n)]}{\partial\boldsymbol{\beta}_n\partial\boldsymbol{\beta}'_n}\right] \tag{6}$$

$$= \sum_{s=1}^{S}\mathbf{X}'_{sn}(\mathbf{P}_{sn} - \mathbf{p}_{sn}\mathbf{p}'_{sn})\mathbf{X}_{sn} - \frac{\partial^2 \log[f(\boldsymbol{\beta}_n)]}{\partial\boldsymbol{\beta}_n\partial\boldsymbol{\beta}'_n} \tag{7}$$

with $L(\boldsymbol{\beta}_n|\mathbf{y}_n^S,\mathbf{X}_n^S)$ the likelihood in (2), $f(\boldsymbol{\beta}_n)$ a prior distribution for individual n's coefficients and $\mathbf{I}_{FIM}(\boldsymbol{\beta}_n,\mathbf{X}_n^S)$ the ordinary Fisher information matrix of the conditional logit choice model. Further, $\mathbf{X}_{sn}$ is the design matrix of choice set s, $\mathbf{P}_{sn} = \mathrm{diag}(p_{1sn},\ldots,p_{Ksn})$ and $\mathbf{p}_{sn} = (p_{1sn},\ldots,p_{Ksn})'$. 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

$$\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n,\mathbf{X}_n^S) = \sum_{s=1}^{S}\mathbf{X}_{sn}'(\mathbf{P}_{sn} - \mathbf{p}_{sn}\mathbf{p}_{sn}')\mathbf{X}_{sn} + \boldsymbol{\Sigma}_0^{-1}. \tag{8}$$

Note that for a uniform prior for instance the second part in (8) will disappear as the prior density is a constant. For a lognormal or exponential prior on the other hand the second part will, just as $\mathbf{I}_{FIM}(\boldsymbol{\beta}_n,\mathbf{X}_n^S)$, involve the coefficients $\boldsymbol{\beta}_n$ which will make the construction of the $\mathcal{D}$-efficient designs more complex. Remark that the first part in (8) will be singular as long as only a few choices have been made by the respondent, the second part ensures that the information matrix is invertible from the start.

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 is assumed and the following criterion is minimized over all possible choice sets to select the first set in the design

$$DB = \int \det\left[\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n,\mathbf{X}_n^1)\right]^{-1/p} f(\boldsymbol{\beta}_n)\, d\boldsymbol{\beta}_n, \tag{9}$$

with $f(\boldsymbol{\beta}_n) \equiv \phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$. Note that the prior initially assumed in (9) may equal the prior used in (5) to construct the Bayesian Fisher information matrix, but this is not necessary.

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

$$f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) = \frac{L(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1},\mathbf{X}_n^{s-1})\phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)}{\int L(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1},\mathbf{X}_n^{s-1})\phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)\, d\boldsymbol{\beta}_n}. \tag{10}$$

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 weighting distribution in the Bayesian $\mathcal{D}$-efficiency criterion to select the next best choice set for respondent n by minimizing

$$DB = \int \det\left[\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n,\mathbf{X}_n^s)\right]^{-1/p} f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n, \tag{11}$$

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.

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. The logarithmic transformation makes the *DB* criterion less sensitive to very small and very large determinant values (Atkinson et al., 2007). This criterion was compared with (11) in Appendix B 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 cfr. Kullback and Leibler (1951)

$$KL(f,g) = E_f\left[\log\frac{f(x)}{g(x)}\right] \tag{12}$$

$$= \int f(x)\log\frac{f(x)}{g(x)}\, dx. \tag{13}$$

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 divergence 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 divergence between the current posterior of the coefficients (obtained with the choice data at hand) and the updated posterior one will 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, 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

$$KLP = \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) KL[f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}), f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})] \tag{14}$$

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 (10). Note that y_{ksn} implies here that the kth alternative would be chosen in choice set s. The weights in (14) are defined as

$$\pi(y_{ksn}|\mathbf{y}_n^{s-1}) = \int p_{ksn}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n \tag{15}$$

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. Remark that the weight in (15) also depends on the choice set $\mathbf{x}_{ksn}$ but this is not mentioned explicitly to simplify the notation.

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 (14) and (15) is replaced by the normal prior $f(\boldsymbol{\beta}_n) \equiv \phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$:

$$\sum_{k=1}^{K} \pi(y_{k1n}) KL[f(\boldsymbol{\beta}_n), f(\boldsymbol{\beta}_n|y_{k1n})] \tag{16}$$

with

$$\pi(y_{k1n}) = \int p_{k1n}(\boldsymbol{\beta}_n) \phi(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)\, d\boldsymbol{\beta}_n. \tag{17}$$

This is in analogy with (9). The same practice is used in all subsequent design algorithms.

Applying the definition of Kullback–Leibler distance from (13), one can show that KLP can be rewritten as

$$KLP = \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) KL[f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}), f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})] \tag{18}$$

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

$$= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int \log \frac{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) \pi(y_{ksn}|\mathbf{y}_n^{s-1})}{f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1})} f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n \tag{20}$$

$$= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int \log \frac{\pi(y_{ksn}|\mathbf{y}_n^{s-1})}{p_{ksn}(\boldsymbol{\beta}_n)} f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n \tag{21}$$

$$= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \left[\log \pi(y_{ksn}|\mathbf{y}_n^{s-1}) - \int \log p_{ksn}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n \right]. \tag{22}$$

From Eq. (22) 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 (11) 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)

$$I_M(X, Y) = \int_Y \int_X f(x, y) \log \frac{f(x, y)}{f(x) f(y)}\, dx\, dy. \tag{23}$$

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. Further, it is clear that this measure is symmetric.

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

$$MUI = \sum_{k=1}^{K} \int f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1}) \log \frac{f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\pi(y_{ksn}|\mathbf{y}_n^{s-1})}\, d\boldsymbol{\beta}_n, \tag{24}$$

or equivalently,

$$MUI = \sum_{k=1}^{K} \left[\int p_{ksn}(\boldsymbol{\beta}_n) \log p_{ksn}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n - \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \log \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \right]. \tag{25}$$

From (25) 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 the expected Kullback–Leibler distance between posteriors.

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

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

$$= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log \frac{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})}\, d\boldsymbol{\beta}_n \tag{28}$$

$$= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) KL[f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}), f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})] \tag{29}$$

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 are very similar but not equal. Yet, due to the high resemblance between both criteria we expect similar results with respect to their design efficiency.

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) and Weissman (2007)

$$H(X) = -\int f(x) \log f(x)\, dx \tag{30}$$

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 the expected posterior entropy (Wang and Chang, 2011) or equivalently maximize

$$ENT = \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\, d\boldsymbol{\beta}_n. \tag{31}$$

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 (31) is equivalent to

$$ENT = \log c + \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log \frac{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})}{c}\, d\boldsymbol{\beta}_n, \tag{32}$$

with c a constant. As $\log c$ can be discarded for maximization, maximizing ENT in (31) equals maximizing the second part in (32), 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

$$MUI = \sum_{k=1}^{K} \int f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1}) \log \frac{f(\boldsymbol{\beta}_n, y_{ksn}|\mathbf{y}_n^{s-1})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\pi(y_{ksn}|\mathbf{y}_n^{s-1})}\, d\boldsymbol{\beta}_n \tag{33}$$

$$= \sum_{k=1}^{K} \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\pi(y_{ksn}|\mathbf{y}_n^{s-1}) \log \frac{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\pi(y_{ksn}|\mathbf{y}_n^{s-1})}{f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\pi(y_{ksn}|\mathbf{y}_n^{s-1})}\, d\boldsymbol{\beta}_n \tag{34}$$

$$= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1})\Big[\int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\, d\boldsymbol{\beta}_n - \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n\Big] \tag{35}$$

$$= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1})[H(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) - H(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})]. \tag{36}$$

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 (Wang and Chang, 2011). 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 (36) 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 estimation and prediction accuracy 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 weighting of the design criteria over priors and posteriors (as for instance in (11), (22) and (25)) 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 research importance sampling is therefore applied to approximate the integrals in this case. More details about importance sampling are given in Appendix C. Note that other approaches, such as normal approximation or Gauss–Hermite quadrature, may be used equally well.

3. Simulation study and results

It has already been elaborately demonstrated in Danthurebandara et al. (2011) and in Yu 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 showing the advantages of an individualized design approach but continues on these findings and considers alternative design criteria, besides $\mathcal{D}$-efficiency, to generate the individual designs. In this section, we compare the *DB*, *KLP*, *MUI* and *ENT* criteria 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, the coding of the levels and the number of alternatives in a choice set. The first scenario assumes that all choice sets comprise two alternatives. 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 $3^3/2/15$. 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 out which criterion is most useful 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 to respectively $(-0.5, 0, -0.5, 0, -0.5, 0)$ and $0.5\mathbf{I}_6$, with $\mathbf{I}_6$ the six-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. Note that for the first scenario also a non-diagonal covariance matrix has been considered, but this did not alter the results.

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 $0.5\mathbf{I}_6$. As also here designs with 15 choice sets are generated, this setup corresponds to $2 \times 3 \times 2 \times 3/3/15$.

In contrast to scenario 1 and scenario 2, in which hypothesized experiments are assumed, the third and the fourth scenario are based on two real transportation studies described in respectively Hess and Polak (2005) and Espino et al. (2008).

Table 1
Overview of the scenarios in the simulation study.

	Experimental setup		
Scenario 1	$3^3/2/15$	$\boldsymbol{\mu}$	$(-0.5, 0, -0.5, 0, -0.5, 0)$
		$\boldsymbol{\Sigma}$	$0.5\mathbf{I}_6$
Scenario 2	$2 \times 3 \times 2 \times 3/3/15$	$\boldsymbol{\mu}$	$(-0.5, -0.5, 0, -0.5, -0.5, 0)$
		$\boldsymbol{\Sigma}$	$0.5\mathbf{I}_6$
Scenario 3	$5 \times 4^2/2/15$	$\boldsymbol{\mu}$	$(-0.068, -0.083, 1.623)$
		$\boldsymbol{\Sigma}$	diag $(0.0009, 0.0015, 0.5622)$
Scenario 4	$3 \times 2 \times 3/3/15$	$\boldsymbol{\mu}$	$(0.419, 0.700, 1.355, 1.638, 1.690)$
		$\boldsymbol{\Sigma}$	diag $(0.0466, 0.9055, 2.6322, 0.5684, 1.1071)$

The experiments deal with air travel choice in the San Francisco bay area (Hess and Polak, 2005) and with the choice of flights connecting the Canary Islands with the Iberian Peninsula (Espino et al., 2008). In scenario 3, three attributes characterize the profiles, more specifically the access-cost, the in-vehicle access-time and the flight frequency. Note that this is a fragment of the attributes considered in Hess and Polak (2005). We assume the first attribute to have five levels, the remaining two four and choice sets include two alternatives. As in Hess and Polak (2005), the attributes are continuously coded which is most frequent in transportation. Choice experiments with 15 sets are designed.

In the final setup also three attributes are considered, but they now describe the type of free food available on board, the comfort (more specifically the leg room) and the reliability of compensation when the flight is delayed or canceled. This represents again a part of the setup in Espino et al. (2008). Respectively three, two and three levels are assumed for these attributes, the choice sets include three alternatives and the designs 15 choice sets. In agreement with the assumptions in Espino et al. (2008), the attributes in scenario 4 are dummy coded. The true values for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for both scenarios can be found in Table 1.

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. Relying on the results from Danthurebandara et al. (2011) and Yu et al. (2012), we assume that the relative performance of the design criteria will not alter when applying distinct prior information and therefore prior miss-specification is not considered in this research.

Further, to sample from the prior and the importance distributions in order to approximate the integrals in the design criteria, extensible shifted lattice points transformed with Baker's transformation are applied instead of random sampling. In case of a model with few parameters, the speed of sampling could be increased by using the spherical–radial transformation method (Yu et al., 2010). In view of extensions to this study, we however rely on Yu et al. (2010) and use lattice points for sampling which is most efficient when the number of parameters is large. For the results presented below, 512 draws are used.

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 estimation errors RMSE_{μ} and RMSE_{Σ}, which are given by

$$\text{RMSE}_{\mu} = \sqrt{\frac{(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})}{p}}, \tag{37}$$

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

$$\text{RMSE}_{\Sigma} = \sqrt{\frac{(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})'(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})}{p_{\sigma}}}, \tag{38}$$

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

Note that for each scenario and for each design criterion, the generation of the choice designs and the estimation of the mixed logit choice model was repeated 100 times. The mean RMSE_{μ} and RMSE_{Σ} values are given in Fig. 1 for each scenario and all four design criteria. We tested the differences between the results obtained with the four design criteria using the test for random block designs, treating the simulations as random blocks and the four criteria as different treatments. No significant differences in RMSE_{μ} and RMSE_{Σ} could be observed, the population parameters in the model are estimated equally accurate with all four design criteria. These results are a first indication that the Kullback–Leibler design criteria are useful

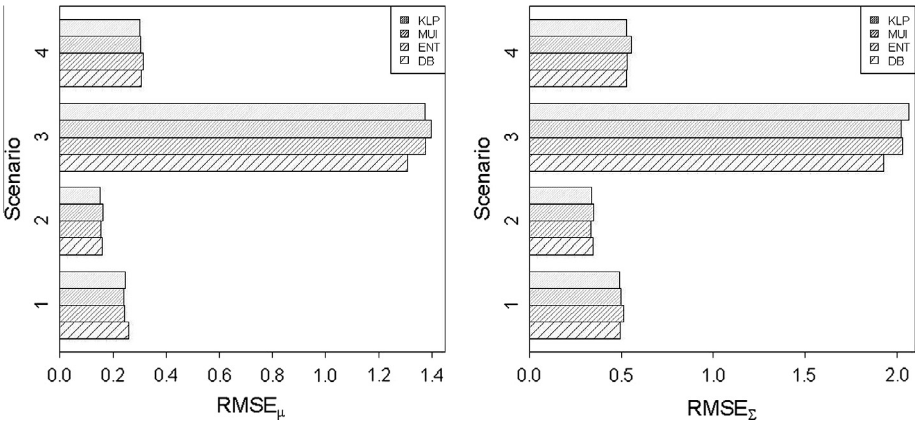

Fig. 1. Mean RMSE$_\mu$ and RMSE$_\Sigma$ values obtained with *KLP*, *MUI*, *ENT* and *DB* for the different scenarios.

alternatives to $\mathcal{D}$-efficiency as design criterion for the mixed logit choice model. The four design methods yield equally accurate estimates for the population parameters.

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 RMSE$_\mu$ and RMSE$_\Sigma$, the root mean squared estimation error RMSE$_\beta$ is also compared between the design criteria:

$$\mathrm{RMSE}_{\boldsymbol{\beta}} = \sqrt{\frac{1}{N}\sum\nolimits_{n=1}^{N} \frac{(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)}{p}}, \tag{39}$$

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. Fig. 2 shows the mean values of RMSE$_\beta$ over the 100 simulations for each design criterion and each scenario. Also with respect to the individual coefficients, no clear differences in estimation error are observed.

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, 3160 and 816. As such, the number of profiles in $\mathbf{p}$ is respectively 21,420, 6320 and 2448. The prediction error is now defined as

$$\mathrm{RMSE}_{\mathbf{p}} = \sqrt{\frac{1}{N}\sum\nolimits_{n=1}^{N} \frac{[\mathbf{p}(\hat{\boldsymbol{\beta}}_n) - \mathbf{p}(\boldsymbol{\beta}_n)]'[\mathbf{p}(\hat{\boldsymbol{\beta}}_n) - \mathbf{p}(\boldsymbol{\beta}_n)]}{SK}}, \tag{40}$$

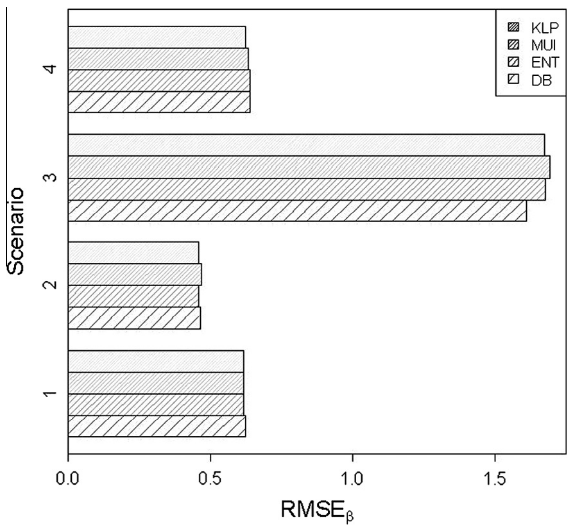

Fig. 2. Mean RMSE$_\beta$ values obtained with *KLP*, *MUI*, *ENT* and *DB* for the different scenarios.

with S the number of choice sets and K the number of alternatives in each set. We tested the differences in prediction error similarly as before but as no significant differences between the design criteria were observed, the values are not reported. Yet, these results confirm the equal performance of the different design criteria.

Finally, in addition to the estimation and prediction errors, we also compare the design criteria with respect to bias in the individual-specific coefficients' estimates. For this, the average bias over individuals is defined in the following way:

$$\mathrm{BIAS} = \sum_{n=1}^{N}\sum_{m=1}^{p}\frac{|\hat{\beta}_{nm} - \beta_{nm}|}{Np}. \tag{41}$$

Table 2 reports the mean BIAS values over the simulation repetitions, standard deviations are given between brackets. It is clear that for none of the scenarios considered significant differences in average bias can be detected between the design criteria.

To examine bias for each parameter in the scenarios separately, the average estimates of the coefficients are plotted against their true values for all individuals. Examples of such plots are given in Fig. 3 for a parameter estimated with and without bias. The first parameter in scenario 1 is for instance clearly biased downwards with all design criteria (left panel of Fig. 3). On the other hand, for the second parameter in scenario 4 no overall bias is present. The small amount of data, both with respect to the sample size and the data available per respondent, probably explains the potential bias. More important in our comparison study however, is that the different design criteria perform equivalently in terms of bias. No criterion causes more bias than the others.

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 perform equivalently.

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 weighting 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 3 displays the average computation time (in seconds) to sequentially generate one additional set in a choice experiment using respectively *KLP*, *MUI*, *ENT* and *DB* as selection rule. The computation times were obtained

Table 2
Mean BIAS values (standard deviations) for *KLP*, *MUI*, *ENT* and *DB*.

Scenario	*KLP*	*MUI*	*ENT*	*DB*
1	0.487 (0.051)	0.489 (0.046)	0.491 (0.055)	0.492 (0.047)
2	0.361 (0.034)	0.369 (0.037)	0.360 (0.037)	0.365 (0.036)
3	1.924 (0.834)	1.915 (0.818)	1.897 (0.802)	1.916 (0.797)
4	0.491 (0.045)	0.499 (0.044)	0.501 (0.047)	0.501 (0.045)

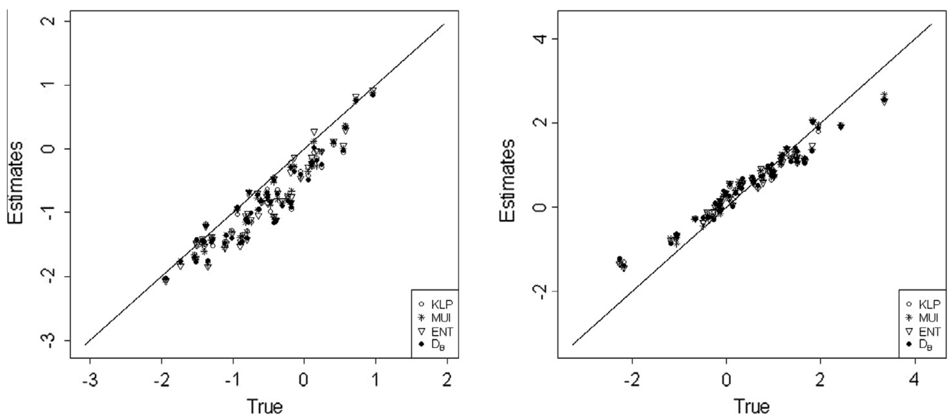

Fig. 3. Average estimates of the individual coefficients against their true values for the first coefficient in scenario 1 (left) and the second coefficient in scenario 4 (right).

Table 3
Average computation time (seconds) to select one additional choice set with *KLP*, *MUI*, *ENT* and *DB* using various numbers of draws.

	512	1024	2048	512	1024	2048
	Scenario 1			Scenario 2		
KLP	0.078	0.154	0.287	1.729	3.401	6.752
MUI	0.087	0.156	0.296	1.786	3.492	6.978
ENT	0.094	0.172	0.334	1.976	3.876	7.691
$\mathcal{D}_B$	1.514	2.986	5.937	31.318	62.090	126.262
	Scenario 3			Scenario 4		
KLP	0.553	1.055	2.084	0.224	0.425	0.825
MUI	0.553	1.073	2.117	0.228	0.436	0.850
ENT	0.624	1.208	2.372	0.258	0.484	0.940
$\mathcal{D}_B$	10.668	21.478	43.339	3.404	6.791	13.493

using a Dell personal computer with a 2.80 GHz Intel Processor and 3.45 GB RAM. Note that besides 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. Note, for instance, that in the second scenario using 512 draws, for which the selection of a choice set takes approximately only 1.8 s with *KLP*, *MUI* and *ENT*, the use of *DB* is far less attractive as each respondent must wait on average more than 31 s for every additional choice set. 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.

$$ENT = \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn}) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}, y_{ksn})\, d\boldsymbol{\beta}_n \tag{42}$$

$$= \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \int \frac{p_{ksn}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})}{\pi(y_{ksn}|\mathbf{y}_n^{s-1})} \log \frac{p_{ksn}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})}{\pi(y_{ksn}|\mathbf{y}_n^{s-1})}\, d\boldsymbol{\beta}_n \tag{43}$$

$$= \sum_{k=1}^{K} \int p_{ksn}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) \log \frac{p_{ksn}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})}{\pi(y_{ksn}|\mathbf{y}_n^{s-1})}\, d\boldsymbol{\beta}_n \tag{44}$$

$$= \sum_{k=1}^{K} \int p_{ksn}(\boldsymbol{\beta}_n) \log p_{ksn}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n - \sum_{k=1}^{K} \pi(y_{ksn}|\mathbf{y}_n^{s-1}) \log \pi(y_{ksn}|\mathbf{y}_n^{s-1}) + \sum_{k=1}^{K} \int p_{ksn}(\boldsymbol{\beta}_n) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n \tag{45}$$

$$= MUI + \sum_{k=1}^{K} \int p_{ksn}(\boldsymbol{\beta}_n) \log f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1}) f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})\, d\boldsymbol{\beta}_n \tag{46}$$

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

Note that, instead of the $\mathcal{D}$-efficiency criterion, one could think of applying $\mathcal{A}$-efficiency to decrease the computation time as the latter only requires to minimize a sum, more specifically the sum of the variances of the parameter estimators, and thus not involves the computation of a determinant. However, you do need to compute the inverse of the Fisher information matrix, which is at least as time consuming. Therefore, no time is gained with the $\mathcal{A}$-efficiency criterion (Kessels et al., 2006).

Although very fast, even for the *KLP* criterion highly complex choice experiments become problematic. As all possible (remaining) choice sets are evaluated when selecting an additional set in the individual designs, the speed of the design process also clearly depends on the total amount of choice sets that can be constructed from the profile attributes and the levels considered. If for instance the profiles in the choice sets are characterized by four attributes with four levels each, already 32,640 different choice sets with two alternatives can be constructed. This is a very large amount and will highly increase the computation time. Shifting to three alternatives in each set the number of possible sets increases to 2,763,520, which makes an evaluation of all sets even with *KLP* infeasible. Therefore, in case of highly complex setups, the Kullback–Leibler criteria can only remain practical by restricting the design space.

3.3. Initial choice sets

To conclude this simulation section, we briefly discuss the potential inclusion of initial sets in the individualized design of choice experiments for the mixed logit choice model (Danthurebandara et al., 2011; Yu et al., 2011). 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 a non-sequential way in advance of collecting data, 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 Sections 2.2.1 and 2.2.2 by iteratively updating the posterior.

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

$$\int \det\left[\mathbf{I}_{BFIM}(\boldsymbol{\beta}_n, \mathbf{X}_n^{S_I})\right]^{-1/p} f(\boldsymbol{\beta}_n)\, d\boldsymbol{\beta}_n, \tag{47}$$

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 (9), 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 (Bliemer and Rose, 2010b). 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 gainful than updating it after each set from the start.

For all four scenarios in Table 1, *KLP*, *MUI*, *ENT* and *DB* designs (15 sets in total) were generated with five initial sets. However, for none of the scenarios considered they appeared beneficial.

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 panel version of the mixed logit choice model appeared only feasible for local optimality criteria (Bliemer and Rose, 2010a), Yu et al. (2011) introduced an individualized design approach applying $\mathcal{D}$-efficiency to estimate the model. This research continues on these findings but focusses 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. Though defined by different concepts, all three criteria can be written as Kullback–Leibler information measures.

A simulation study illustrates that the efficiency of the designs obtained with the different criteria is equivalent. Both the population parameters and the individual-specific coefficients of the mixed logit choice model are estimated equally accurate with choice data from the four types of efficient designs. Note however that we did not investigate the impact of mis-specified priors in this paper but relied on previous studies that showed that the relative performance of design criteria for mixed logit choice models can be derived from correctly specified prior information. Remark further that only the statistical efficiency of the designs is considered in the study. This means that we evaluate the accuracy of the parameter estimates obtained with the different individual designs, but that we do not discuss the designs' empirical efficiency. Many attributes used in transportation studies have a clear rank order of preference in their levels, which means that choice sets with a dominant alternative can be constructed. Although such choice sets do not give any additional information about travelers' preferences, as one may assume that each rational respondent will always choose the dominant alternative, these sets might be efficient from a statistical point of view and might therefore pop up in the statistically efficient designs (Crabbe and Vandebroek, 2012). We disregard this dominance issue in this research and do not take the empirical efficiency of the individual designs into account when comparing the different design criteria.

The Kullback–Leibler design criteria are however preferred over the traditional $\mathcal{D}$-efficiency criterion to set up transport choice experiments 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 18 times longer with the *DB* criterion than with the Kullback–Leibler criteria. In a research environment evolving to the use of online, interactive choice experiments, speed is not only a huge merit but even a necessity. Where individualized design was the solution for efficiently designing choice experiments for the mixed logit choice model, the Kullback–Leibler criteria warrant the feasibility. The criteria are unfortunately also subjected to limitations. As discussed in Section 3.2, highly increasing the complexity of the experimental setup (choice sets with many alternatives and many high-leveled attributes characterizing the alternatives), makes even the Kullback–Leibler criteria impracticable. Therefore further research for speed improvements remains necessary.

Acknowledgement

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

Appendix A. Algorithm to generate individualized designs for the mixed logit choice model with *KLP*, *MUI*, *ENT* and *DB*

1. Assume an initial prior for the individual-specific coefficients $\boldsymbol{\beta}_n$, for instance a normal prior $\mathcal{N}(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ for which the values $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are common for all individuals.
2. Efficiently construct the initial choice set $\mathbf{X}_n^1$ for individual n by optimizing either the *KLP*, *MUI*, *ENT* or *DB* criterion with the prior $\mathcal{N}(\boldsymbol{\beta}_n|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$.
3. Observe individual n's choice $\mathbf{y}_n^1$ for the initial choice set $\mathbf{X}_n^1$ and update the prior information on the partworths $\boldsymbol{\beta}_n$ in a Bayesian way with the observed choice data $\mathbf{y}_n^1$:

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

4. Select the subsequent choice set for individual n by optimizing *KLP*, *MUI*, *ENT* or *DB*, with $f(\boldsymbol{\beta}_n|\mathbf{y}_n^1)$ as weighting distribution for $\boldsymbol{\beta}_n$ to compute the criteria.
5. Repeat steps 3 and 4 until a specific amount of choice sets is reached. More specifically, assume individual n has completed $s-1$ choice sets and $\mathbf{X}_n^{s-1}$ and $\mathbf{y}_n^{s-1}$ denote respectively the choice design and the observed choice data for these sets. The updated posterior information on $\boldsymbol{\beta}_n$ then equals

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

This updated posterior is now used as the weighting distribution in the Bayesian design criteria to select the next best choice set s for respondent n (see Eqs. (11), (22), (25) and (46)).

Appendix B. 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:

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

To compare the estimation accuracy of both criteria, the density curves of the $\text{RMSE}_{\boldsymbol{\beta}}$ values for scenario 1 are shown in Fig. B.4, the plots are similar for the remaining scenarios. The graphs illustrate that the *DB* and the $DB_{\log}$ criterion are equivalent in design efficiency and this was confirmed by a paired sample t-test. Similar plots have been obtained for $\text{RMSE}_{\boldsymbol{\mu}}$ and $\text{RMSE}_{\boldsymbol{\Sigma}}$.

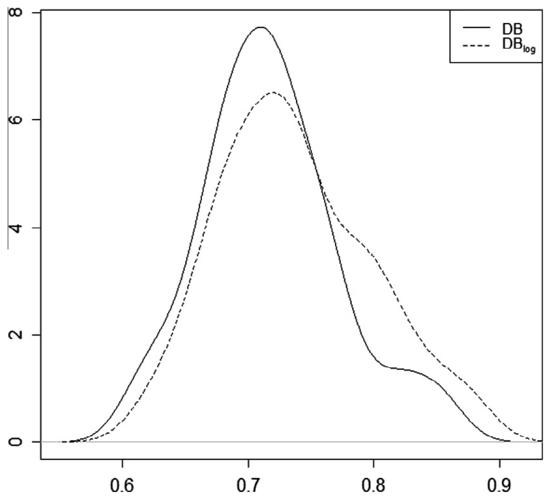

Fig. B.4. Density curves of the $\text{RMSE}_{\boldsymbol{\beta}}$ values obtained with *DB* and $DB_{\log}$ for scenario 1.

Appendix C. Importance sampling

Consider a measure C, depending on β, averaged over a weighting distribution $f(\beta)$ for β, i.e.:

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

As in most cases such an integral is analytically very complex, approximation techniques are used to solve it numerically. When $f(\beta)$ has a known and closed form, e.g. normal, draws β^r can be easily obtained and applied to approximate the measure above by an equally weighted sum over these draws:

$$\frac{1}{R}\sum_{r=1}^{R} C(\beta^r).$$

Sometimes however, the weighting distribution $f(\beta)$ does not have a closed form, which makes sampling from it infeasible. This is for instance the case for the updated posteriors $f(\boldsymbol{\beta}_n|\mathbf{y}_n^{s-1})$ considered in this research as their normalizing constant is unknown. To circumvent this problem, importance sampling is one of the techniques that can be applied then to approximate the integral above.

Instead of sampling from the distribution of interest $f(\beta)$, an importance density $g(\beta)$ is chosen that approximates $f(\beta)$ and from which draws can easily be obtained. With draws β^r from $g(\beta)$ the expression above is approximated by

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

with w_r the importance weight attached to draw β^r from the importance density. The weights are defined by the kernels f^* and g^* of the distributions f and g in the following way (Albert, 2009):

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

This makes sense as $\int f(\beta)\, d\beta = 1$ and therefore

$$\int C(\beta)\, f(\beta)\, d\beta = \frac{\int C(\beta) \frac{f(\beta)}{g(\beta)}\, g(\beta)\, d\beta}{\int \frac{f(\beta)}{g(\beta)}\, g(\beta)\, d\beta}.$$

Approximating the integrals in both the numerator and the denominator with the same draws from $g(\beta)$, we get the expression for the importance sampling approximation.

A convenient and frequently chosen importance density for a posterior is the t density. More specifically, for the updated posteriors in this research, we assume as importance density 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 (Albert, 2009; Yu et al., 2011).

References

Albert, J., 2009. Bayesian Computation with R. Springer, New York.

Atkinson, A., Donev, A., Tobias, R., 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 (3), 173–185.

Ben-Akiva, D., Bolduc, D., Bradley, M., 1993. Estimation of travel choice models with randomly distributed values of time. Transportation Research Record 1413, 88–97.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer Series in Statistics. Springer-Verlag.

Bhat, C.R., 1998. Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. Transportation Research Part A 32 (7), 495–507.

Bhat, C.R., 2012. Recent developments in discrete choice model formulation, estimation and inference. Transportation Research Part B 46 (2), 273–275.

Bhat, C.R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. Transportation Research Part B 46 (7), 817–833.

Bliemer, M.C.J., Rose, J.M., 2010a. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transportation Research Part B 44 (6), 720–734.

Bliemer, M.C.J., Rose, J.M., 2010b. Serial choice conjoint analysis for estimating discrete choice models. In: Hess, S., Daly, A. (Eds.), Choice Modelling: The State-of-the-Art and the State-of-the-Practice. Emerald, Bingley, UK, pp. 139–161.

Chang, H.H., Ying, Z., 1996. A global information approach to computerized adaptive testing. Applied Psychological Measurement 20 (3), 213–229.

Cheng, Y., 2009. When cognitive diagnosis meets computerized adaptive testing: CD-CAT. Psychometrika 74 (4), 619–632.

Crabbe, M., Vandebroek, M., 2012. Using appropriate prior information to eliminate choice sets with a dominant alternative from $\mathcal{D}$-efficient designs. Journal of Choice Modelling 5 (1), 22–45.

Danthurebandara, V., Yu, J., Vandebroek, M., 2011. Sequential choice designs to estimate the heterogeneity distribution of willingness-to-pay. Quantitative Marketing and Economics 9 (4), 429–448.

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

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 (4), 593–606.

Glickman, M.E., Jensen, S.T., 2005. Adaptive paired comparison design. Journal of Statistical Planning and Inference 127 (1–2), 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 (1), 75–92.
Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. Transportation 30 (2), 133–176.
Hess, S., Hensher, D.A., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. Transportation Research Part B 44 (6), 781–790.
Hess, S., Polak, J.W., 2005. Accounting for random taste heterogeneity in airport-choice modelling. Transportation Research Record 1915, 36–43.
Hess, S., Train, K., 2011. Recovery of inter- and intra-personal heterogeneity using mixed logit models. Transportation Research Part B 45 (7), 973–990.
Hess, S., Bierlaire, M., Polak, J.W., 2005. Estimation of value of travel time savings using mixed logit models. Transportation Research Part A 39 (2-3), 221–236.
Hess, S., Adler, T., Polak, J.W., 2007. Modelling airport and airline choice behaviour with the use of stated preference survey data. Transportation Research Part E 43 (3), 221–233.
Hess, S., Smith, C., Falzarano, S., Stubits, J., 2008. Measuring the effects of different experimental designs and survey administration methods using an Atlanta managed lanes stated preference survey. Transportation Research Record 2049, 144–152.
Jourdan, A., Franco, J., 2010. Optimal latin hypercube designs for the Kullback–Leibler criterion. Advances in Statistical Analysis 94 (4), 341–351.
Kessels, R., Goos, P., Vandebroek, M., 2006. A comparison of criteria to design efficient choice experiments. Journal of Marketing Research 43 (3), 409–419.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. The Annals of Mathematical Statistics 22 (1), 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 (2), 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. Springer, New York, pp. 77–101.
Revelt, D., Train, K., 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. Review of Economics and Statistics 80 (4), 647–657.
Rose, J.M., Bliemer, M.C.J., 2009. Constructing efficient stated choice experimental designs. Transport Reviews 29 (5), 587–617.
Sándor, Z., Wedel, M., 2002. Profile construction in experimental choice designs for mixed logit models. Marketing Science 21 (4), 455–475.
Toubia, O., Hauser, J.R., Simester, D.I., 2004. Polyhedral methods for adaptive choice-based conjoint analysis. Journal of Marketing Research 41 (1), 116–131.
Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press.
Wang, C., Chang, H.H., 2011. Item selection in multidimensional computerized adaptive testing – gaining information from different angles. Psychometrika 76 (3), 363–384.
Weissman, A., 2007. Mutual information item selection in adaptive classification testing. Educational and Psychological Measurement 67 (1), 41–58.
Yu, J., Goos, P., Vandebroek, M., 2009. Efficient conjoint choice designs in the presence of respondent heterogeneity. Marketing Science 28 (1), 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 (8), 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 (4), 378–388.
Yu, J., Goos, P., Vandebroek, M., 2012. A comparison of different Bayesian design criteria for setting up stated preference studies. Transportation Research Part B 46 (7), 789–807.