Effect of choice complexity on design efficiency in conjoint choice experiments

Vishva Danthurebandara, Jie Yu and Martina Vandebroek

Effect of choice complexity on design efficiency in conjoint choice experiments

Vishva Manohara Danthurebandara

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

Email: Vishva.Danthurebandara@econ.kuleuven.be

Tel: +32 (0)16 32 69 63

Fax: +32 (0)16 32 66 24

Jie Yu

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

Email: Jie.Yu@econ.kuleuven.be

Tel: +32 (0)16 32 69 62

Fax: +32 (0)16 32 66 24

Martina Vandebroek

Faculty of Business and Economics & Leuven Statistics Research Centre, Katholieke
Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.

Email: Martina.Vandebroek@econ.kuleuven.be

Tel: +32 (0)16 32 69 75

Fax: +32 (0)16 32 66 24

Effect of choice complexity on design efficiency in conjoint choice experiments

Abstract

Conjoint choice experiments have become a powerful tool to explore individual preferences. The consistency of respondents' choices depends on the choice complexity. For example, it is easier to make a choice between two alternatives with few attributes than between five alternatives with several attributes. In the latter case it will be much harder to choose the preferred alternative which is reflected in a higher response error. Several authors have dealt with this choice complexity in the estimation stage but very little attention has been paid to set up designs that take this complexity into account. The core issue of this paper is to find out whether it is worthwhile to take this complexity into account in the design stage. We construct efficient semi-Bayesian D-optimal designs for the heteroscedastic conditional logit model which is used to model the across respondent variability that occurs due to the choice complexity. The degree of complexity is measured by the Entropy, as suggested by Swait and Adamowics (2001). The proposed designs are compared with a semi-Bayesian D-optimal design constructed without taking the complexity into account. The simulation study shows that it is much better to take the choice complexity into account when constructing conjoint choice experiments.

1 Introduction

Conjoint choice experiments have become increasingly popular for collecting and studying preferences of individuals. Discrete choice models are usually derived under the assumption of utility-maximizing behavior of the decision makers. In a random utility model, the utility that a decision maker obtains from an alternative is described by a structural part, with information about the alternative, and an error term which represents all other influences. In most applications, this error term is assumed to have the same variance in all choice sets. However, according to Swait and Adamowics (2001), people use different strategies to cope with complex situations. Therefore one can expect that the error variance will vary with the complexity of the choice set. In this paper we will use the heteroscedastic model that was proposed to model this between respondent variability and develop optimal designs to estimate this model efficiently.

In the literature of conjoint choice experiments, there is only limited research on how changes in the structure of the choice set changes choice outcomes (DeShazo and Fermo, 2002). The complexity of a choice set is however crucial, because it directly affects the choice consistency. That it is easier to make a choice between two alternatives with few attributes than between five alternatives with a lot of attributes will be reflected in a higher response error in the latter case. In the literature, several measurements have been introduced to quantify the choice complexity. DeShazo and Fermo (2002) used five measurements, which describe the structure of the choice set. Sándor and Franses (2009) used two of the complexity measures of DeShazo and Fermo (2002) and one price related measure, which is the key factor of their empirical study. Severin (2000) used one major complexity measurement which is the number of trade-offs that respondents have to make in his or her decision process. This can also be referred to as the similarity of alternatives in terms of attribute levels. Mazzotta and Opaluch (1995) and Dellaert et al. (1999) use the number of attributes that vary across the alternatives to measure the complexity in their research on choice consistency and complexity. In all the studies mentioned above, statistics that describe the structure of the choice set are used to measure the complexity. Swait and Adamowicz (2001) made an argument that each of these measurements is a component of complexity rather than an overall measure. Therefore, they introduced entropy as an overall complexity measure, which summarizes the impact of the number of alternatives, the number of attributes, the correlation among attributes and the similarity among utilities of alternatives.

The statistical design is one of the key challenges in implementing a conjoint experiment, since the efficiency of the parameter estimates depends on the design. Most of the authors referred to above do not assess the effect of choice complexity on the statistical design of the experiment.

By ignoring choice complexity when designing the experiment, the choice data obtained will be inconsistent with the estimation model. Hence, the experimental design obtained cannot be optimal.

The core issue we address in our paper is to investigate whether it is worthwhile to take the choice complexity into account when we construct the design. Standard models assume that respondents have unlimited information processing capacity, which allows them to make their choice in a strictly optimal way irrespective to the complexity of the choice situation (de Palma et al., 1994). The heteroscedastic conditional logit model proposed by Swait and Adamowics (2001) however uses the scale factor to bring the complexity into the model. We use their parameterization to model the between respondent variability that occurs due to the choice complexity. In our research, we propose efficient semi-Bayesian D-optimal designs, constructed by considering the choice complexity. The proposed designs are compared with two semi-Bayesian D-optimal designs which are constructed without considering the choice complexity but for the rest uses the same design setting as the proposed design.

We organize this paper as follows. In section 2, we present the theoretical model, discuss the complexity measure, the design efficiency criterion, the design construction algorithm and the benchmark designs we used. This is followed in section 3 by a relative design efficiency study. In section 4, we present the simulation study setup, the proposed design settings and the estimation results. We discuss the impact of the misspecification of the complexity function in section 5 and finally in section 6 we evaluate and summarize our key findings.

2　Methodology

2.1　Heteroscedastic Conditional Logit Model

First consider the homoscedastic conditional logit model (McFadden, 1974), which is popular for analyzing the data from conjoint choice experiments. The random utility a given respondent n attaches to an alternative k in choice set s is given as

$$U_{ksn} = \mathbf{x}'_{ks}\boldsymbol{\beta} + \varepsilon_{ksn} \tag{1}$$

where, $\mathbf{x}_{ks}$ is a p-dimensional vector containing the attribute values of alternative k in choice set s, $\boldsymbol{\beta}$ is a p-dimensional vector of parameters and U_{ksn} is the utility that the decision maker n

actually obtained from alternative k in choice set s. The error term ε_{ksn} is assumed to have an extreme value distribution. Assuming there are K alternatives in a choice set, the probability that alternative k is chosen from choice set s is

$$q_{ks} = \frac{exp\left(\mu\, \mathbf{x}'_{ks}\boldsymbol{\beta}\right)}{\sum_{i=1}^{K} exp\left(\mu\, \mathbf{x}'_{is}\boldsymbol{\beta}\right)}, \qquad k = 1, ..., K \qquad (2)$$

where μ is the scale factor. The scale factor is defined as $\frac{\pi}{\sqrt{6}\sigma}$, where σ is the standard error of ε. This model is called the homoscedastic conditional logit model since the scale factor is assumed constant.

An increase in choice set complexity will add noise to the error term of the random utility function (DeShazo and Fermo 2002) which is reflected in a higher response error σ. Thus, the error variance σ^2 depends on the choice complexity and we will denote this dependency explicitly in

$$p_{ks} = \frac{exp\left(\mu(C_s)\, \mathbf{x}'_{ks}\boldsymbol{\beta}\right)}{\sum_{i=1}^{K} exp\left(\mu(C_s)\, \mathbf{x}'_{is}\boldsymbol{\beta}\right)} \qquad (3)$$

where C_s measures the complexity of choice set s. The model is called the heteroscedastic conditional logit model (HCLM) by DeShazo and Fermo (2002) and Swait and Adamowics (2001).

When the choice situation becomes more complex, respondents use different methods to make their decision process simple which generates higher choice inconsistencies across respondents. So we assume that tastes, represented by $\boldsymbol{\beta}$, are homogeneous across individuals but introduce the possibility that the error term is heteroscedastic across individuals due to complexity. In other words, this model can test whether the variability between respondents can be explained by some complexity measures. This model for between respondent variability, was also used by Swait and Adamowics (2001) and Amaya-Amaya et al. (forthcoming).

Instead of modeling heterogeneity across individuals some researchers have used the HCLM to model the individual variation in the error term from choice set to choice set. For example, Sándor and Franses (2009), DeShazo and Fermo (2002) use HCLM to model the consistency of an individual's behavior.

2.2 Measure of Choice Complexity

To describe choice complexity, most authors use measures which describe the structure of the choice set. In this subsection we will discuss them in detail and define the measure of complexity that we apply in this paper.

DeShazo and Fermo (2002) use five complexity measures: the number of alternatives, the number of attributes, the variance of the attribute levels across alternatives, the mean standard deviation of attribute levels for each alternative in the choice set and the dispersion of the standard deviation of each alternative in the choice set. Sándor and Franses (2009) also use the mean standard deviation of attribute levels for each alternative in the choice set in addition to the number of trade offs.

Swait and Adamowics (2001) use a single measurement to assess the choice complexity, namely the *entropy*. Entropy is a measure of the uncertainty associated with a random variable. For example, a fair coin flip (2 equally likely outcomes) will have less entropy than a roll of a die (6 equally likely outcomes). By definition, entropy measures the uncertainty of a situation and is therefore an indication of complexity. That is, higher complexity or higher uncertainty is reflected by higher entropy. This can be explained in terms of choice probabilities. Figure 1 is a plot of choice probability against entropy. In Figure 1, the X-axis indicates the choice probability of alternative 1. All other alternatives are assumed to have equal probability and the probability of alternative 1 varies from 0 to 1. Consider the case of 2 alternatives (red line). When the choice probability of the first alternative is zero, there is no complexity at all in the choice set since the second choice set has maximum choice probability. This is reflected by zero entropy. When the choice probability of the alternative 1 is equal to 1, again there is no complexity in the choice set since the alternative 1 dominates the other. Entropy is maximized when the choice probability of the first alternative is equal to 0.5. This is where the choice set is more complex since both alternatives have the same probability to be chosen. In the case of 3 alternatives, the highest entropy comes when the choice probabilities are equal to each other (0.33). As entropy is a proper summary measurement of many other complexity measures used in the literature (see Swait and Adamowics, 2001) we will use entropy as the measure of choice complexity. Entropy is defined as

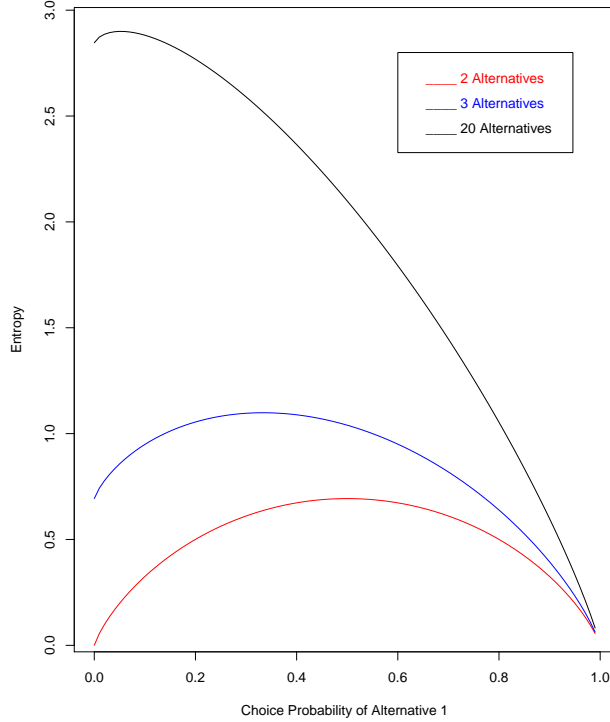$$H_s = -\sum_{k=1}^{K} q_{ks} \ln q_{ks} \tag{4}$$

Figure 1: Entropy vs. Choice Probability for different choice set sizes

where q_{ks} as defined in (2).

As we described earlier, the variance of the error term of the random utility function is affected by the complexity of the choice situation. Therefore, we use the scale factor which is inversely proportional to the error variance to bring the complexity into the model. In the literature, we can find two ways to relate the scale factor to the complexity measure. Sándor and Franses (2009) use an exponentiated linear function of the complexity measures as the scale function. As they model the inconsistency in individual choices, they assume that the accuracy of the decision process monotonically decreases when the complexity increases.

Swait and Adamowics (2001) use psychological insights to derive that respondents will apply more effort to make consistent decisions up to a certain point of complexity, but after that level they simplify their decision process by applying simple theories and using prior knowledge which generates greater inconsistencies across respondents. So when the choice complexity is low (i.e. when the choice situation has dominant alternatives), the variance is also low since the choice of best alternative is easy to select with low effort. In medium complex situations, variance is high since different respondents use different heuristics and different effort levels which will add more noise across respondents. When the complexity is high (i.e. equal choice probabilities), variance decreases compared to the medium complexity situation because all

respondents will choose one of the alternatives with equal utility regardless their effort level. These intuitive explanations suggest that the scale, or decision accuracy, has an U-shape behavior with the choice set complexity.

To capture this nonlinearity, the scale factor is assumed to be an exponentiated quadratic function of the entropy:

$$\mu\left(C_s\right) = exp\left(\theta_1 H_s + \theta_2 H_s^2\right). \tag{5}$$

The exponential function ensures non negativity and also convergence properties (DeShazo and Fermo, 2002). The U-shape is obtained by restricting θ_1 and θ_2, i.e., $\theta_1 \leqslant 0$ and $\theta_2 \geqslant 0$.

2.3 Design Efficiency Criteria

The log likelihood function of the heteroscedastic conditional logit model can be written as

$$LL(y|\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{n=1}^{N}\sum_{s=1}^{S}\sum_{k=1}^{K} y_{ksn} \ln p_{ksn} \tag{6}$$

where y_{ksn} is 1 if decision maker n chooses alternative k in choice set s and 0 otherwise, and p_{ksn} is the probability that y_{ksn} is equal to 1.

Considering all the merits of the D-criterion in this context (see Yu et al., 2008), we choose this criterion in our study. The D-criterion is based on the Fisher's Information matrix and minimizes the determinant of the inverse information matrix. The information matrix $I(\boldsymbol{\beta}, \boldsymbol{\theta}|X)$ of the heteroscedastic conditional logit model can be derived by taking the second order derivatives of the log-likelihood function with respect to both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. It is a $(p+2) \times (p+2)$ dimensional matrix where p is the number of parameters corresponding to the attribute levels and 2 is the number of parameters in the complexity measure (θ_1 and θ_2). The expression of this information matrix is rather complicated and can be found in appendix.

The design efficiency measure corresponding to the D-criterion is the D-error, which is based on the determinant of the information matrix. As this determinant depends on the parameter values, we consider the expected value of the D-error over some prior distribution π_0:

$$D_{SB}^C - error = \int_{\mathfrak{R}^{p+2}} det[I(\boldsymbol{\beta}, \boldsymbol{\theta}|X)]^{-\frac{1}{p+2}} \pi_0(\boldsymbol{\beta}, \boldsymbol{\theta}) \, d(\boldsymbol{\beta})d(\boldsymbol{\theta}) \tag{7}$$

which is called the semi-Bayesian D-error. The design with the smallest $D^C_{SB} - error$ is called the optimal design and is denoted by X^C_{SB}.

In practice, this semi-Bayesian D-error has to be approximated:

$$\widetilde{D}^C_{SB} - error \simeq \frac{1}{R} \sum_{r=1}^{R} det[I_{FIM}((\boldsymbol{\beta},\boldsymbol{\theta})^r|X)]^{-\frac{1}{p+2}} \tag{8}$$

where R is the total number of random draws taken from the prior distribution $\pi_0(\boldsymbol{\beta},\boldsymbol{\theta})$. Following Yu et al. (2008), we assume a multivariate normal prior distribution and take R=2000 random draws from the prior distribution to approximate the $D^C_{SB} - error$.

2.4 Design Construction Algorithm

In this study, we use the coordinate-exchange algorithm proposed by Meyer and Nachtsheim (1995) which is a computationally efficient algorithm. It starts with a random design and replaces only one coordinate or attribute level of a profile at each step. A level change is accepted if and only if it results in a better $D^C_{SB} - error$. After the first iteration, that is, when the algorithm has found the best exchange for all attributes of all profiles of the design, the algorithm goes back to the first attribute of the first profile and continues until no significant improvement is possible anymore. To avoid poor local optima, we used 1500 runs of the algorithm to find the designs reported in this paper.

2.5 Benchmark Designs

The X^C_{SB} designs that we derive for the heteroscedastic conditional logit model taking into account the choice complexity are compared with some benchmark designs. These benchmark designs are also semi-Bayesian D-optimal designs but optimized for the homoscedastic conditional logit model(see Sándor and Wedel, 2001 and Kessels et al., 2006). These designs are denoted by X_{SB}.

3 Relative Design Efficiency (RDE)

In this section we will compare the efficiency of the proposed designs X_{SB}^C with the efficiency of the benchmark designs X_{SB}. To measure how well a X_{SB}^C design performs compared to another design we use the relative efficiency. The relative design efficiency of design X_{SB} relative to design X_{SB}^C is defined as

$$RDE = \frac{D_{SB}^C - error(X_{SB}^C)}{D_{SB}^C - error(X_{SB})}. \tag{9}$$

Note that for values below 1, the design X_{SB} is less efficient for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ than the design X_{SB}^C.

We consider 15 choice sets, each choice set having 2 alternatives and the alternatives described by 4 attributes with 3 levels. This design setting is denoted by $3^4/2/15$. We used effect coding to assign [1 0], [0 1], [-1 -1] to the levels 1, 2, 3 respectively. The prior distribution for $\boldsymbol{\beta}$ is multivariate normal with mean $\boldsymbol{\mu}_{0\beta} = [-1, 0, -1, 0, -1, 0, -1, 0]$ and variance covariance matrix $\mathbf{I_p}$. Specifying such a mean value for $\boldsymbol{\beta}$ ensures that the mean parameters corresponding to each of the attribute levels are evenly spaced between -1 and 1. For each three level attribute, the values -1, 0 and 1 then correspond to the attribute levels 1, 2 and 3. Thus it is assumed that the utility increases with the level of each attribute (Yu et al., 2008 and Kessels et al., 2006). The prior distribution for $\boldsymbol{\theta}$ is multivariate normal with mean $\boldsymbol{\mu}_{0\theta} = [1, -1]$ and variance covariance matrix $0.1\,\mathbf{I_2}$ and is assumed to be independent from the prior of $\boldsymbol{\beta}$. The benchmark design X_{SB} has the same specification and uses the same prior for $\boldsymbol{\beta}$.

We evaluate the efficiency of the two designs in 88 different parameter spaces. For each parameter space, the parameters are drawn from a multivariate normal distribution with mean $\begin{pmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\theta}} \end{pmatrix}$ and covariance matrix $\sum$ where,

$$\begin{pmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\theta}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_{0\beta} \\ \boldsymbol{\mu}_{0\theta} \end{pmatrix} + \gamma\,\mathbf{1_{p+2}}$$

$$\sum = \sum\nolimits_0 + \delta\,\mathbf{I_{p+2}}$$

and $\sum_0 = \text{diag}(1,...,1,\ 0.1,\ 0.1)$ is a $(p+2) \times (p+2)$ diagonal matrix, $\mathbf{1_{p+2}}$ and $\mathbf{I_{p+2}}$ denote a $(p+2) \times 1$ vector of ones and a $(p+2) \times (p+2)$ identity matrix respectively. Note that $\boldsymbol{\mu}_{0\beta}$, $\boldsymbol{\mu}_{0\theta}$

and $\sum_0$ are the parameter values we used to construct the design X_{SB}^C. The mean deviation factor γ takes values in the interval [-0.5, 0.5] and the variance deviation factor δ takes values in the interval [-0.2, 0.5]. The parameter γ reflects how much the mean of the multivariate normal distribution that we use to assess the $D_{SB}^C - error$ deviates from the mean value assumed in the design stage. The parameter δ reflects how much the variances differ from the ones assumed in the design stage.
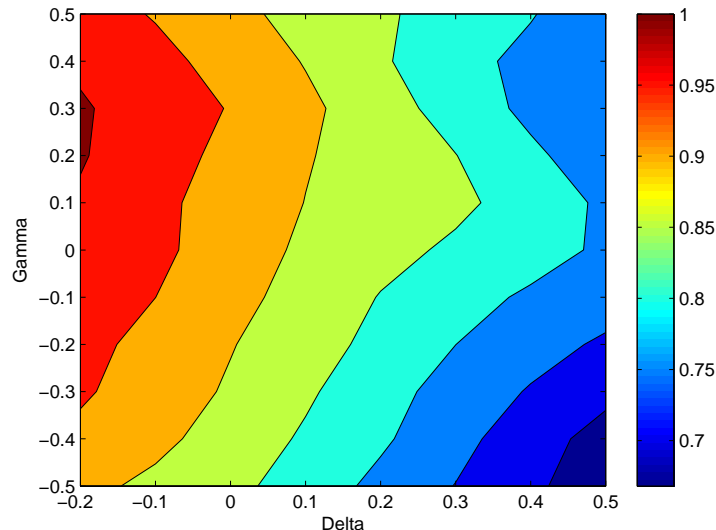

Figure 2: Relative design efficiency of the proposed design over the benchmark design

The relative design efficiency values for each of these 88 parameter spaces are represented in Figure 2 and are divided into several categories. The plot shows that the proposed design is more efficient than the benchmark design within the whole parameter space we considered. Only the red area shows scenarios where both designs have comparable efficiencies.

Figure 2 can be used to assess the sensitivity of the designs to the misspecification of the prior. The parameter space defined by $\gamma = 0$ and $\delta = 0$ is the space we used to construct the designs. It can be seen that the proposed design is quite robust against the misspecification of the mean of the prior distribution but less robust with respect to the misspecification of the variances in the prior distribution.

4 Simulation Study

4.1 Designs Considered in the Simulation

We construct X_{SB}^C designs for different prior distributions and different experimental setups. An overview of these designs is given in Table 1. The first four designs have 15 choice sets, each choice set having 2 alternatives and each alternative containing 4 attributes with 3 levels. Each of these designs is constructed using the same prior distribution for $\boldsymbol{\beta}$ but with different prior distributions for $\boldsymbol{\theta}$. The prior distribution for $\boldsymbol{\beta}$ is multivariate normal with mean $\boldsymbol{\mu}_{0\beta} = [-1, 0, -1, 0, -1, 0, -1, 0]$ and variance covariance matrix $\mathbf{I_p}$. We used four different multivariate normal prior distributions for $\boldsymbol{\theta}$; $\boldsymbol{\theta} \sim N(\boldsymbol{\mu}_{0\theta}, \sigma_{0\theta}^2 \mathbf{I_2})$ where $\boldsymbol{\mu}_{0\theta}$ takes the values [-1, 1.7] or [-2, 3] and $\sigma_{0\theta}$ can be 0.1 or 0.5. These prior values are obtained from the empirical study of Swait and Adamowics (2001). These prior values also ensure that the scale as a function of complexity is U-shaped. We compare these 4 designs with the semi-Bayesian D-optimal design X_{SB} for the corresponding homoscedastic model using the same prior for $\boldsymbol{\beta}$.

The last four designs have a smaller number of parameters but three alternatives per choice set. Each alternative is described using two attributes each having three levels and the design has only 6 choice sets. In this case the benchmark design is again a semi-Bayesian D-optimal design, X_{SB}, constructed without considering the choice complexity. Similar to the first four designs, we keep the same prior distribution for the $\boldsymbol{\beta}$, fixed at mean $\boldsymbol{\mu}_{0\beta} = [-1, 0, -1, 0]$ and variance $\sigma_{0\beta}^2 \mathbf{I_p}$.

Table 1: Overview of the designs used in the simulation study

		Prior Values			
Design	Specification	$\beta \sim N(\boldsymbol{\mu}_{0\beta}, \mathbf{I_p})$	$\theta \sim N(\boldsymbol{\mu}_{0\theta}, \sigma_{0\theta}^2 \mathbf{I_2})$		Benchmark
		$\boldsymbol{\mu}_{0\beta}$	$\boldsymbol{\mu}_{0\theta}$	$\sigma_{0\theta}$	
1	$3^4/2/15$	(-1 0 -1 0 -1 0 -1 0)	(-1 1.7)	0.1	Semi-Bayesian D-optimal Design
2	$3^4/2/15$	(-1 0 -1 0 -1 0 -1 0)	(-1 1.7)	0.5	with the same specification and priors
3	$3^4/2/15$	(-1 0 -1 0 -1 0 -1 0)	(-2 3)	0.1	but generated without considering
4	$3^4/2/15$	(-1 0 -1 0 -1 0 -1 0)	(-2 3)	0.5	the choice complexity (X_{SB})
5	$3^2/3/6$	(-1 0 -1 0)	(-1 1.7)	0.1	Semi-Bayesian D-optimal Design
6	$3^2/3/6$	(-1 0 -1 0)	(-1 1.7)	0.5	with the same design specification
7	$3^2/3/6$	(-1 0 -1 0)	(-2 3)	0.1	and priors but without taking the
8	$3^2/3/6$	(-1 0 -1 0)	(-2 3)	0.5	choice complexity into account (X_{SB})

4.2 Simulation Setup

For each design setting in Table 1, we drew 120 parameter values ($\boldsymbol{\beta}$, $\boldsymbol{\theta}$) from the prior distribution. For each draw, we simulated responses for 100 respondents based on the X_{SB}^C and corresponding X_{SB} design and estimated the parameters. The Expected Root Mean Square Error of Estimation (ERMSE) value is then calculated as explained in the next section. Finally, we average all the ERMSE values over the 120 true parameter values and use this average value to represent the estimation performance of a design.

4.3 Expected Root Mean Square Error of Estimation ($\boldsymbol{ERMSE}$)

To assess how well parameters can be estimated using the different designs, we compute the *ERMSE*. This measure compares how close the estimated parameters are to the true parameters:

$$ERMSE(\boldsymbol{\lambda}) = \int_{\mathfrak{R}^q} [(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})'(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})]^{1/2}\, \varphi(\widehat{\boldsymbol{\lambda}})\, d(\widehat{\boldsymbol{\lambda}}) \tag{10}$$

where λ is the parameter vector we consider to calculate the ERMSE. For example, if we calculate the ERMSE for the total parameter vector then $\boldsymbol{\lambda} = (\boldsymbol{\beta}, \boldsymbol{\theta})'$ and q = p+2. If we calculate the ERMSE for $\boldsymbol{\beta}$ alone then $\boldsymbol{\lambda} = \boldsymbol{\beta}$ and q = p. Also note that, $\varphi(\widehat{\boldsymbol{\lambda}})$ is the distribution of the estimates, $\widehat{\boldsymbol{\lambda}}$ is the vector of estimated parameters and $\boldsymbol{\lambda}$ is the vector of true parameters.

In practice, ERMSE values will be approximated by

$$ERMSE(\widetilde{\boldsymbol{\lambda}}) \simeq \frac{1}{M} \sum_{m=1}^{M} [(\widehat{\boldsymbol{\lambda}}^m - \boldsymbol{\lambda})'(\widehat{\boldsymbol{\lambda}}^m - \boldsymbol{\lambda})]^{1/2} \tag{11}$$

where M is the number of simulations and $\widehat{\boldsymbol{\lambda}}^m$ is the parameter estimate obtained from the mth simulated data set.

To compare the estimation accuracy of different designs, we compute the percentage decrease in *ERMSE*. This measure is defined by

$$1 - \frac{ERMSE(X_{SB}^C)}{ERMSE(X_{SB})}. \tag{12}$$

Positive values of this measure show that estimation accuracy increases by using the X_{SB}^C design compared to the benchmark design X_{SB}.

Table 2 shows the $ERMSE(\boldsymbol{\beta}, \boldsymbol{\theta})$ values and the percentage decrease in $ERMSE(\boldsymbol{\beta}, \boldsymbol{\theta})$ by using the proposed design instead of the benchmark design. Note that the $ERMSE$ is computed by taking into account both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

Table 2: Evaluation of the estimation accuracy in terms of $ERMSE(\boldsymbol{\beta}, \boldsymbol{\theta})$ and percentage decrease of $ERMSE(\boldsymbol{\beta}, \boldsymbol{\theta})$

Design	$ERMSE(\boldsymbol{\beta}, \boldsymbol{\theta})$		Percentage Decrease
	Proposed	Benchmark	
1	1.1748	1.5907	26%
2	1.2183	1.5172	20%
3	1.2917	1.6778	23%
4	1.3362	1.6766	20%
5	0.3925	0.5644	30%
6	0.5552	0.8579	35%
7	0.4204	0.6610	36%
8	0.6334	0.9524	33%

As can be seen in Table 2, designs constructed by considering the choice complexity, perform better in terms of estimation accuracy for all designs we looked at. On average, the decrease in $ERMSE$ is more than 25% which is quite considerable.

To investigate whether this improvement comes from $\boldsymbol{\beta}$ or from $\boldsymbol{\theta}$ or from both of them, we compute the $ERMSE(\boldsymbol{\beta})$ and $ERMSE(\boldsymbol{\theta})$ separately. Table 3 presents the results of $ERMSE(\boldsymbol{\beta})$ and $ERMSE(\boldsymbol{\theta})$ for the last four designs with three alternatives. It can be seen that the proposed designs are, on average, 35% more accurate than the benchmark design in terms of $\boldsymbol{\beta}$. This is a very encouraging result as $\boldsymbol{\beta}$ is the parameter vector that people are most interested in. Also note that, similar results were obtained for the other designs.

To better visualize the accuracy of the estimates, we plot the distribution of the estimated parameter values. We sampled 120 vectors of true parameters and for each true parameter vector, we simulated data with both designs X_{SB}^{C} and X_{SB} and estimated the parameters. Figure 3 shows the distribution of the estimated $\boldsymbol{\beta}$ parameters for five randomly selected true parameter vectors. In Figure 3, the black dashed line represents the estimation density obtained by the design X_{SB}^{C} and the red line represents the benchmark design X_{SB} for design number 5 in Table 1. Moreover, the vertical line shows the true parameter in each case.

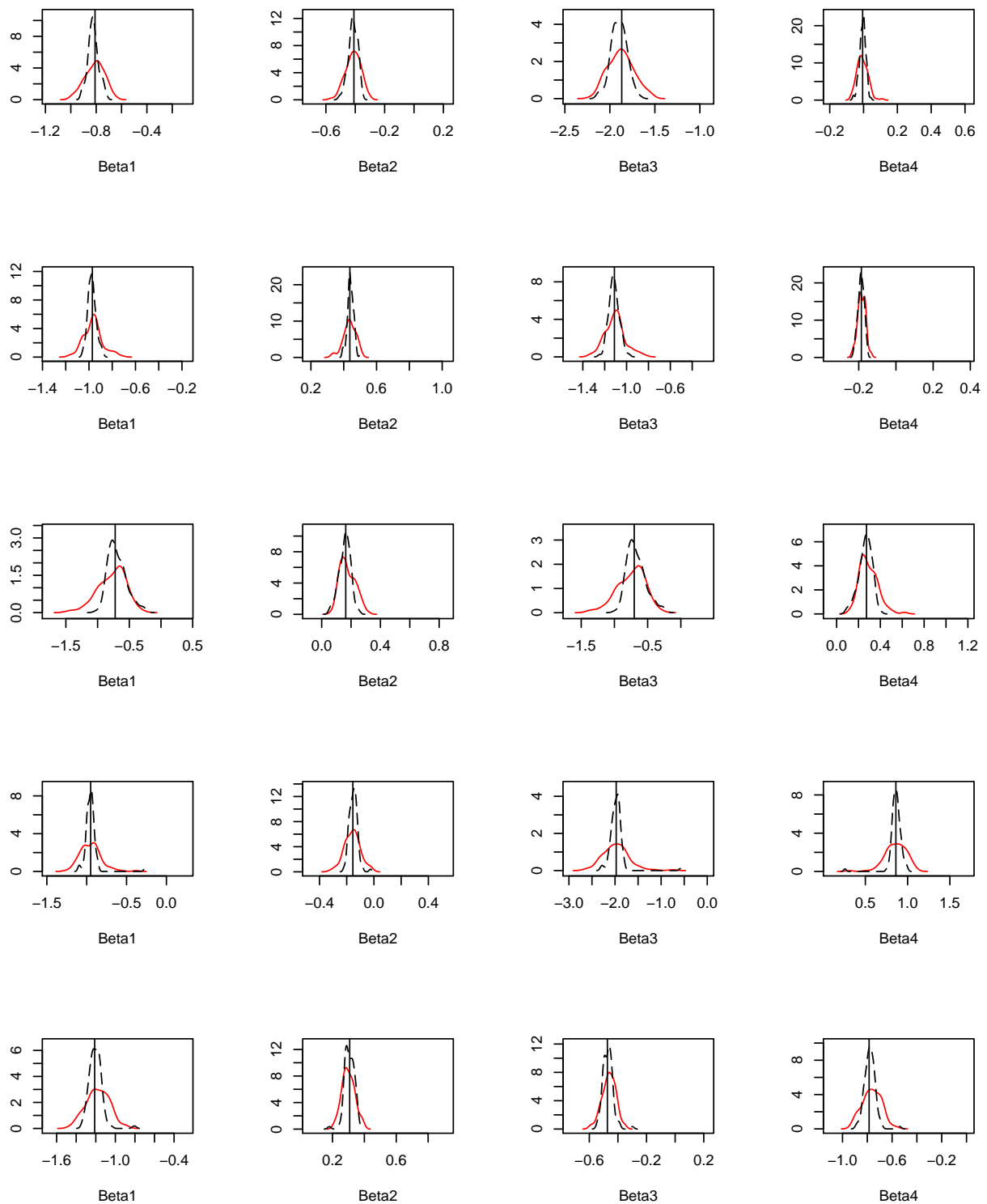As can be seen in Figure 3, in most of the cases the proposed design performs much better than the benchmark.

Figure 3: Distributions of estimated $\boldsymbol{\beta}$ parameters under X_{SB}^{C} and X_{SB} designs - Design $3^2/3/6$ (Design number 5 in Table 1)

Table 3: Evaluation of the estimation accuracy in terms of $ERMSE(\boldsymbol{\beta})$ and $ERMSE(\boldsymbol{\theta})$ seperately and percentage decrease of $ERMSE$

| Design | Parameter set | ERMSE | | Percentage Decrease |
		Proposed	Benchmark	
5	Total	0.3925	0.5644	30%
	$\boldsymbol{\beta}$	0.1254	0.2070	39%
	$\boldsymbol{\theta}$	0.3537	0.4711	25%
6	Total	0.5552	0.8579	35%
	$\boldsymbol{\beta}$	0.1506	0.2530	40%
	$\boldsymbol{\theta}$	0.5174	0.7773	33%
7	Total	0.4204	0.6610	36%
	$\boldsymbol{\beta}$	0.1060	0.1627	35%
	$\boldsymbol{\theta}$	0.3910	0.5663	31%
8	Total	0.6334	0.9524	33%
	$\boldsymbol{\beta}$	0.1238	0.1901	35%
	$\boldsymbol{\theta}$	0.6012	0.8264	27%

5 Misspecification of the Scale Function

Parameterizations of the complexity function in the HCLM as used by Swait and Adamowics (2001) assures that the variance as function of the choice set complexity is inversely U-shaped. This shape was found in their empirical study when analyzing some data sets. But they also obtained constantly decreasing patterns of variance with increasing complexity for some of the data sets they analyzed. This pattern can be modeled by a exponentiated linear complexity function.

In this section we assess the robustness of the designs against misspecification of the complexity function. We study how much we will loose by using a exponentiated quadratic complexity function to construct the design when there is no quadratic effect in reality and how much we will loose by using a exponentiated linear complexity function to construct the design when there is a quadratic effect.

To investigate this, we constructed two semi-Bayesian D-optimal designs what we call the quadratic design and the linear design, using the following two specifications of the scale factor, respectively:

$$\mu\left(C_s\right) = exp\left(\theta_1 H_s + \theta_2 H_s^2\right)$$

and

$$\mu\left(C_s\right) = exp\left(\theta_1 H_s\right).$$

Note that the design setting is $3^2/3/6$. The prior distributions of the $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are assumed to be independent multivariate normal with the following parameters.

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_{0\beta}, \mathbf{I_p}) \text{ where } \boldsymbol{\mu}_{0\beta} = [-1, 0, -1, 0]$$
$$\boldsymbol{\theta} \sim N(\boldsymbol{\mu}_{1\theta}, 0.1\,\mathbf{I_2}) \text{ where } \boldsymbol{\mu}_{1\theta} = [-1, 1.7] \quad \text{(for the quadratic complexity function)},$$
$$\boldsymbol{\theta} \sim N(\boldsymbol{\mu}_{2\theta}, 0.1\,\mathbf{I_2}) \text{ where } \boldsymbol{\mu}_{2\theta} = [-1] \quad \text{(for the linear complexity function)},$$

In the modeling stage, for cases in which no or small quadratic effects are present, we use the quadratic design and the linear design to fit a linear model and compare the resulting estimation performance. This is because in practice, if the quadratic effect is not significant in the modeling stage, the researcher often turns to the linear model. In this way we can investigate how much is lost in terms of estimation performance by using the quadratic design instead of the linear design. For cases in which higher quadratic effects are present in the scale function, we compare the estimation performance obtained from using the quadratic design for fitting a quadratic model with the estimation performance of the linear design for fitting a linear model.

The comparison in terms of estimation performance is visualized in Figure 4 for different sizes of the quadratic effects. The magnitudes of the quadratic effects, parameterized by θ_2, are shown on the x-axis. The estimation performance of the particular model, reflecting by the ERMSE, is shown on the y-axis. Here we compute the ERMSE values only in terms of $\boldsymbol{\beta}$. Note that $\theta_2 = 0$ corresponds to the situation where the model that generates the responses contains no quadratic effects.

The estimation performance of the quadratic design that is used to fit either the quadratic model or the linear model, is visualized by the line labeled "QD". The estimation performance of the linear design that is used to fit the linear model is visualized by the line labeled "LD". The curve "QD" was drawn in the following way: when θ_2 is small, the values shown in the curve QD are the ERMSE($\boldsymbol{\beta}$) values obtained using the quadratic design to fit a linear model. This is because fitting a linear model results in more accurate estimates than fitting a quadratic model. For higher values of θ_2, the curve QD shows the ERMSE($\boldsymbol{\beta}$) values obtained for the quadratic model.

Figure 4 indicates that, when no or small quadratic effects are present, the loss in estimation accuracy by using the quadratic design instead of the linear design to fit a linear model is small, on average 16%. On the other hand, the loss of estimation accuracy by using the linear design rather than the quadratic design, when there are higher quadratic effects in reality, is
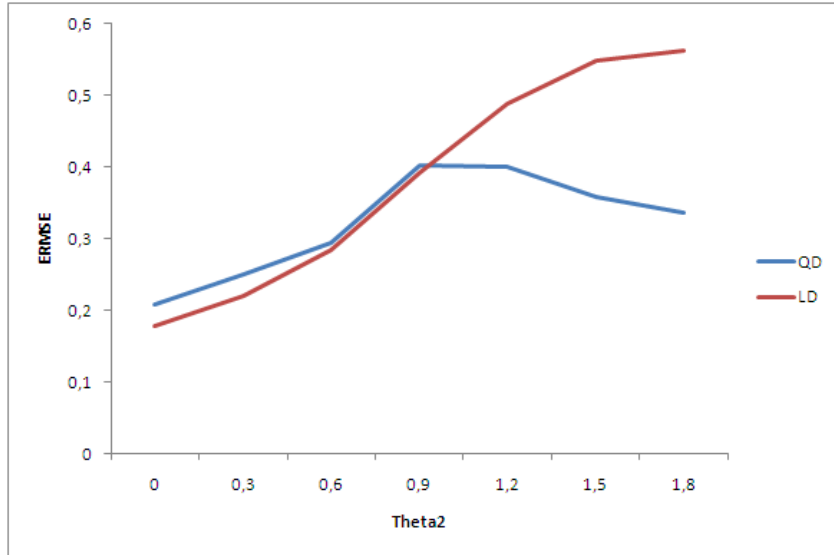
17

Figure 4: Evaluation of estimation accuracy in terms of $ERMSE(\boldsymbol{\beta})$ under different design settings and data generating models

relatively larger. Thus it is clear that, if there is a quadratic effect in reality, then the quadratic complexity function is necessary to obtain accurate estimates. Even if the true complexity function is a linear function of entropy, by using a quadratic complexity function to construct the design one will not loose much. Therefore, it is advised to use a quadratic complexity function in all situations.

6   Discussion and Conclusions

Several authors have addressed the effect of choice complexity at the estimation stage. We examined whether it is worthwhile to take the choice complexity into account at the design stage. We used entropy as a general measure of complexity. The heteroscedastic conditional logit model was used to bring the choice complexity into the model.

In terms of relative design efficiency the proposed designs for the heteroscedastic conditional logit model performed much better than the design which is optimal for the related homoscedastic conditional logit model.

In the simulation study, we constructed 8 semi-Bayesian D-optimal designs considering the choice complexity and compared them with semi-Bayesian D-optimal designs which have the same design settings but are constructed without considering the choice complexity. We found that the average gain in estimation accuracy by using the proposed designs was 25%. Consid-

ering only the improvement in the estimation of $\boldsymbol{\beta}$, an average gain of 35% was obtained.

We also studied the effect of misspecification of the scale factor as a function of the complexity. Based on the simulation results it was concluded that the use of a quadratic complexity function is necessary if there is quadratic effect in reality. By considering the relatively minor loss of estimation accuracy by using a quadratic complexity function to estimate a linear model, it can be conclude that it is best to use the quadratic complexity function in all situations.

Appendix: Derivation of the information matrix

In this appendix we present the derivation of the Fisher information matrix for the heteroscedastic conditional logit model, following Sándor and Franses (2009). We derive it only for one respondent and one choice set. The information matrix is obtained by taking the sum over all respondents and choice sets.

The log likelihood function for one choice set with K alternatives is

$$LL(y|\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{k=1}^{K} y_k \ln p_k = \mathbf{y}' \ln \mathbf{p}$$

where y_k is 1 if decision maker chooses alternative k and 0 otherwise, and p_k is the probability that $y_k = 1$, $\mathbf{y} = (y_1, ..., y_K)'$ and $\mathbf{p} = (p_1, ..., p_K)'$. The Fisher information matrix can be obtained by taking the second order derivatives with respect to both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. The information matrix can be written as

$$I_{FIM}(\boldsymbol{\beta}, \boldsymbol{\theta}|X) = \begin{pmatrix} E[\frac{\partial L}{\partial \beta} \frac{\partial L}{\partial \beta'}] & E[\frac{\partial L}{\partial \beta} \frac{\partial L}{\partial \theta'}] \\ E[\frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \beta'}] & E[\frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \theta'}] \end{pmatrix}.$$

Note that for a parameter vector $\boldsymbol{\tau}$ that is either $\boldsymbol{\beta}$ or $\boldsymbol{\theta}$

$$\frac{\partial L}{\partial \boldsymbol{\tau}} = \left(\mathbf{y}' \frac{\partial \ln \mathbf{p}}{\partial \boldsymbol{\tau}} \right)' = \left(\frac{\partial \mathbf{p}}{\partial \boldsymbol{\tau}'} \right)' \mathbf{P}^{-1} \mathbf{y} \tag{13}$$

where $\mathbf{P} = diag(p_1, ..., p_K)$. The choice probability p_k of the heteroscedastic conditional logit model is

$$p_k = \frac{exp\left[exp\left(\theta_1 H + \theta_2 H^2\right) \mathbf{x}_k' \boldsymbol{\beta}\right]}{\sum_{i=1}^{K} exp\left[exp\left(\theta_1 H + \theta_2 H^2\right) \mathbf{x}_i' \boldsymbol{\beta}\right]}.$$

Then for $\boldsymbol{\tau} = \boldsymbol{\beta}$ and $\boldsymbol{\tau} = \boldsymbol{\theta}$ we obtain

$$\frac{\partial p}{\partial \boldsymbol{\beta}'} = C[\mathbf{I} + \omega \boldsymbol{\beta} \mathbf{A}'][\mathbf{P} - \mathbf{p}\mathbf{p}']\mathbf{X} \tag{14}$$

$$\frac{\partial p}{\partial \boldsymbol{\theta}'} = C[\mathbf{P} - \mathbf{p}\mathbf{p}']\mathbf{X}\boldsymbol{\beta}\mathbf{h} \tag{15}$$

where the design matrix X for one choice set with p attributes and K alternatives is define as

$$X = [X_1 \; X_2 \; .... \; X_p] = \begin{bmatrix} x_{11} & x_{21} & ... & x_{p1} \\ x_{12} & x_{22} & ... & x_{p2} \\ . & . & . & . \\ x_{1K} & x_{2K} & ... & x_{pK} \end{bmatrix}$$

and

$$C = exp\left(\theta_1 H + \theta_2 H^2\right)$$
$$\omega = (\theta_1 + 2\theta_2 H)$$
$$\mathbf{A} = \left(-\sum_{k=1}^{K}[(\ln q_k + 1)(x_{k_1} q_k - q_k \sum_{i=1}^{K} q_i x_{i_1})] \quad .... \quad -\sum_{k=1}^{K}[(\ln q_k + 1)(x_{k_p} q_k - q_k \sum_{i=1}^{K} q_i x_{i_p})] \right)$$
$$\mathbf{h} = \left(H \quad H^2 \right)$$
$$q_k = \frac{exp\left(\mu \, \mathbf{x'_k}\boldsymbol{\beta}\right)}{\sum_j exp\left(\mu \, \mathbf{x'_j}\boldsymbol{\beta}\right)}$$

By using (14) and (15) in the equation 13, we obtain

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = C\mathbf{X'}[\mathbf{P} - \mathbf{pp'}]'[\mathbf{I} + \omega\boldsymbol{\beta}\mathbf{A'}]' \, \mathbf{P^{-1}y},$$

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = C\mathbf{h'}\boldsymbol{\beta'}\mathbf{X'}[\mathbf{P} - \mathbf{pp'}]' \, \mathbf{P^{-1}y}.$$

These derivations can be used to compute the components of the information matrix. We introduce the notations

$$\mathbf{M} = (\mathbf{P} - \mathbf{pp'})\mathbf{X} \qquad \text{and} \qquad \mathbf{B} = (\mathbf{I} + \omega\boldsymbol{\beta}\mathbf{A'})$$

and use the fact that $E[\mathbf{yy'}] = \mathbf{P}$.

$$E[\tfrac{\partial L}{\partial \beta} \tfrac{\partial L}{\partial \beta'}] = C^2 \, \mathbf{M'B'} \, \mathbf{P^{-1}}E[\mathbf{yy'}] \, \mathbf{P^{-1}BM} = C^2 \, \mathbf{M'B'} \, \mathbf{P^{-1}} \, \mathbf{BM}$$

The other components can be computed similarly. The resulting information matrix is

$$I_{FIM}(\boldsymbol{\beta}, \boldsymbol{\theta}|X) = C^2 \begin{pmatrix} \mathbf{M'B'} \, \mathbf{P^{-1}} \, \mathbf{BM} & \mathbf{M'B'P^{-1}M}\boldsymbol{\beta}\mathbf{h} \\ \mathbf{h'}\boldsymbol{\beta'}\mathbf{M'P^{-1}BM} & \mathbf{h'}\boldsymbol{\beta'}\mathbf{M'P^{-1}M}\boldsymbol{\beta}\mathbf{h} \end{pmatrix}.$$

References

Amaya-Amaya, M., Ryan, M., (forthcoming). Between contribution and confusion: an investigation of the impact of complexity in stated preferences choice experiments. *Journal of Health Economics.*

Dellaert, B.G.C., Brazell, J.D., Louviere, J.J., 1999. The effect of attribute variation on consumer consistency. *Marketing Letters.* 10 139-147.

de Palma, A., Myers, G. M., Papageorgiou, Y. Y., 1994. Rational choice under an imperfect ability to choose. *American Economic Review.* 84 419-440.

DeShazo, J., Fermo, G., 2002. Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management.* 44 123-143.

Keller, K.L., Staelin, R., 1987. Effects of quality and quantity of information on decision and effectiveness. *Journal of Consumer Research.* 14 200-213.

Kessels, R., Goos, P., Vandebroek, M., 2006. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research.* 43 409-419.

Meyer, R.K., Nachtsheim, C.J., 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics.* 37 60-69.

Mazzotta, M., Opaluch, J., 1995. Decision making when choices are complex: A test of Heiner's hypothesis. *Land Economics.* 71(4) 500-515.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In Frontiers in Econometrics, ed. Paul Zarembka: 105-142. New York: Academic Press.

Sándor, Z., Franses, P.H., 2009. Consumer price evaluations through choice experiments. *Journal of Applied Econometrics.* 24 517-535.

Sándor, Z., Wedel, M., 2001. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research.* 38 430-444.

Severin, V.C., 2000. Comparing statistical efficiency and respondent efficiency in choice experiments. The University of Sydney.

Swait, J., Adamowicz, W., 2001. Choice environment, market complexity and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes.* 86 141-167.

Vermeulen, B., Goos, P., Vandebroek, M., 2008. Models and optimal designs for conjoint choice experiments including a no-choice option. *International Journal of research in Marketing.* 25 94-103.

Yu, J., Goos, P., Vandebroek, M., 2008. Model robust design of conjoint choice experiments. *Journal of Communications in Statistics - Simulation and Computation.* 37 1603-1621.