# Experimental Investigation of Consumer Price Evaluations

Zsolt Sándor[1]

Erasmus University Rotterdam

and

Philip Hans Franses

Erasmus University Rotterdam

ECONOMETRIC INSTITUTE REPORT 2004-12

[1]Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: sandor@few.eur.nl.

# Experimental Investigation of Consumer Price Evaluations

**Abstract**

We develop a procedure to collect experimental choice data for estimating consumer preferences with a special focus on consumer price evaluations. For this purpose we employ a heteroskedastic mixed logit model that measures the effect of the way prices are specified on the variance of choice. Our procedure is based on optimal design ideas from the statistics literature and on some algorithms for constructing choice designs published in marketing journals.

In an empirical application on mobile phone preferences we find evidence that the way prices are specified significantly affects the variance of choice. In a simulation study we show that our design is significantly more efficient than randomly generated designs, which can be regarded as equivalent to most commonly used experimental designs in the literature.

JEL No.: C25, C81, C9.

Key words: demand, heterogeneity, task complexity, Bayesian design, quasi-random.

# 1 Introduction

In numerous situations when real-life data are not available researchers use experimental data. Such situations may occur, for example, when one would like to estimate, or test hypotheses about, consumer preferences. In several disciplines of economics and related fields choice experiments are conducted and the obtained data are analyzed through discrete choice models. The recent literature has witnessed a growing interest for such choice experiments in various demand studies, for example, on environmental issues (e.g., Adamowicz et al. 1997, Layton and Brown 2000), on transportation problems (e.g., Brownstone and Train 1999, Hensher and Sullivan 2003), on health care issues (e.g., Scott 2001, San-Miguel et al. 2002), in marketing (where the literature is huge; we only mention a pioneering work by Louviere and Woodworth 1983) and on other demand problems (Revelt and Train 1998, van Ophem et al. 1999).

In many of these studies observations from several thousands of consumers were collected in the experiments, which incurred large costs for conducting the experiments. However, in many of the above studies, by carefully designing the experiment, it is possible to reduce the number of respondents significantly and still obtain the same amount of information on the quantities of interest. This has been done for experiments in biology, physics, chemistry, etc., and it has led to the optimal design literature in statistics, where these techniques have been investigated for several decades. Starting in the 1990's a number of papers published in marketing journals considered the problem of optimal design of choice experiments (e.g., Huber and Zwerina 1996, Arora and Huber 2001, Sándor and Wedel 2001), and this line of research gradually adopted more and more advanced techniques from the statistics literature. For example, Sándor and Wedel

(2003) demonstrate that their procedure produces designs, which, although not optimal, have highly improved efficiency compared to its predecessors.

Despite the need for designs with improved efficiency and despite the advances in constructing such designs, there is hardly any empirical study that considers efficiency aspects for designing a choice experiment concerning consumer preferences. Some of the marketing papers mentioned above contain empirical analyses but these are focused on illustrating specific features in the proposed design procedures and hence fail to study important practical aspects that these imply. We attempt to fill this gap in the present paper. More precisely, we design an experiment based on efficiency considerations, in which we adopt a number of ideas from the optimal design literature in statistics combined with some advances in the experimental choice design literature in marketing. We discuss in detail the problems that occur during the process of design construction for commonly used discrete choice models. Some of these problems are not trivial and not yet encountered in the literature.

The designs we construct have two novel features. The first is that we construct so-called Bayesian designs (i.e., designs based on the prior distribution of the parameters) for a version of the mixed (i.e., random coefficient) logit model (McFadden and Train 2000). This involves a large computational burden in terms of Monte Carlo simulations, which we solve by employing so-called quasi-random samples (e.g., Niederreiter 1992) instead of the pseudo-random samples used in Monte Carlo simulations. The second novel feature is that we construct our design with improved efficiency so that it accounts for choice task complexity in the design. This aspect is important because explicit modelling of the effect of the experimental setup on the outcome of the experiment

enables the experiment to better approximate the real-life preferences. In this respect we rely on recent results by Swait and Adamowicz (2001) and DeShazo and Fermo (2002), who made important steps in measuring the effects of the experimental setup on the overall precision of the estimates.

In this paper we design a choice experiment and collect data for estimating preferences for mobile phones with a special focus on measuring the effect of specifying the price. This problem is motivated by the question whether consumers are able to consciously evaluate prices in the presence of much distracting information, like price discounts and other characteristics of the product. We provide more details about this topic in Section 2, where in addition we put forward our model, and discuss how we incorporate the choice complexity. In Section 3, we describe how we construct the experimental design. In Section 4, we discuss specific issues regarding the collection of the data and present the estimation results on mobile phone preferences. We conclude the paper with a section containing a summary and various possible topics for further research.

# 2 The model

We use a mixed logit model modified so that it accounts for choice task complexity. The mixed logit model has been successfully employed in the econometrics and related literature (e.g., Allenby and Lenk 1994, Berry et al. 1995, Brownstone and Train 1999) to model consumer heterogeneity, which we believe is relevant for our investigation as well. The effect of task complexity on the variance of choice is an important aspect

in choice experiments. A choice task that is more complex is expected to have a less precise choice response. If task complexity is significant and not captured by the model then it will yield less precise estimates of the parameters of interest.

In a typical choice experiment a respondent faces the task of choosing the best from one or more sets of hypothetical products. Suppose that each respondent is given $S$ choice sets, each consisting of $J$ hypothetical products, to choose the most preferred from each. We define the utility for a respondent $i$ who chooses hypothetical product $j$ from choice set $s$ as a heteroskedastic mixed logit, that is,

$$u_{ijs} = x'_{ijs}\left(\beta + V_{is}\sigma\right) + \sigma_{is}\varepsilon_{ijs},$$

where $x_{ijs}$ is a $K \times 1$-vector of characteristics of product $j$ from choice set $s$, $\beta$ is the mean and $\sigma$ is the standard deviation parameter of the random coefficient of $x_{ijs}$, $V_{is}$ is a diagonal matrix with standard normal random variables on its diagonal, which represent latent consumer characteristics, and $\varepsilon_{ijs}$ is an iid type I extreme value random variable that is common to logit models. Since it has a coefficient $\sigma_{is}$, the variance of the utility depends on the choice set. Throughout this paper we sometimes refer to $\sigma_{is}^2$ as the variance of choice.

Note that respondents are assumed to have different latent characteristics $V_{is}$ in different choice sets. There are two reasons for adopting this assumption. The first is that the implied model is more general than the one that assumes the same latent characteristics $V_i$ over the choice sets, and there are reasons to believe that respondents may treat different choice tasks differently (Louviere 2003). We note that some authors use the

model with the same latent characteristics over the choice sets (e.g., Revelt and Train 1998). The second reason for adopting the model with different latent characteristics is that it simplifies the computation of the information matrix needed in the design construction.[2]

For incorporating the effect of task complexity on the variance of choice we rely on Swait and Adamowicz (2001) and DeShazo and Fermo (2002). In fact, in our formulation from the previous paragraph we follow the latter paper. We define $\sigma_{is}$ so that it depends on the characteristics of the hypothetical products and the price specification in choice set $s$ as well as some parameters, that is, $\sigma_{is} = \exp(-c_{is}'\gamma)$, where $c_{is}$ is a column vector characterizing the complexity of the choice task in choice set $s$. The parameter vector $\gamma$ then measures the effect of task complexity on the variance of choice. We give more details on the vector $c_s$ below in section 2.2 when discussing the experiment.

## 2.1 Price evaluations

Much of the literature on consumer choice assumes that consumers are rational utility maximizers. One reason for this assumption is that it allows for the application of reasonably easy to analyze models for observed consumer choice. In contrast, there is also substantial literature on consumer decision making which allows for the possibility that consumers do not always behave perfectly rational, even in case they intend to do so, see Bettman et al. (1993). This phenomenon is often coined as bounded rationality, see Rubinstein (1998). The drivers of this bounded rationality are found in the effort

[2] We may view this as a situation where we have two rival models, and prior to estimation we are not able to decide which represents our data. It is possible to take such model uncertainty into account when constructing the design; see for example Ponce De Leon and Atkinson (1991) and Montepiedra and Yeh (1998).

that consumers have to exercise to make a choice. This effort depends on the difficulty of the choice task, and hence it can depend on the number of alternatives, on the way the choice alternatives are specified, on the number of product characteristics involved and also on other factors (e.g., Tyebjee 1979, Johnson and Payne 1985).

In this paper we examine whether it can happen that choice complexity leads to making mistakes about prices of products. Price is just one of the many possible features of a product, and it may well be that other product information may distract consumers from paying explicit attention to only the price.

One may now wonder whether consumers are able to consciously evaluate prices, in the presence of much distracting information. For example, and as we will consider below, can consumers understand that 135 Euro is the same price as 150 Euro with a 10 per cent discount, in case the product also has a variety of other features? Our experimental investigation below will illustrate that consumers do indeed face difficulties here.

## 2.2 Task complexity variables

In the model we use two variables for measuring choice task complexity. For specifying the first variable we rely on DeShazo and Fermo (2002). We define this variable for choice sets with two alternatives, since we only use such choice sets. For choice set $s$ we define

$$c_{s1} = \frac{SD_1 + SD_2}{2}, \text{ where for alternative } j, \; SD_j = \sqrt{\frac{1}{K}\sum_{k=1}^{K} i_k \left(z_{jk} - \overline{z}_j\right)^2}. \qquad (1)$$

Here $i_k$ is 0 if $z_{1k} = z_{2k}$ and 1 otherwise, that is, it cancels the effect on $c_{s1}$ of those characteristics for which the levels are the same. We note that in our study we have characteristics with two and three levels (see Table A2 in Appendix A) and $z_{jk}$ is the characteristic $k$ of alternative $j$ recoded so that it takes -1, 0, 1 for each three-level characteristic and -0.5, 0.5 for each two-level characteristic, increasing as the attractiveness of the characteristic increases (e.g., for the highest price it is -1 and for the lowest price it is 1).

The variable $SD_j$ captures the dispersion within an alternative excluding the characteristics with the same levels in the two alternatives. If the dispersions for the alternatives are large then so will $c_{s1}$ be. For example, if a choice set contains hypothetical products that have either only highly attractive or only highly unattractive characteristics then $c_{s1}$ will be small reflecting that the task of choosing from this choice set is easy. We note that $c_{s1}$ is a combination of two variables defined by DeShazo and Fermo, that is, the average dispersion and the number of characteristics whose levels differ across alternatives. These authors also included the standard deviation of the dispersions as another variable that can potentially affect choice task complexity. Although they find this variable significant in their empirical analysis, its effect on the variance of choice turns out to be minor, so we do not include it in our model.

We define our second task complexity variable $c_{s2}$ as the variable by which we measure the effect of price evaluation on the variance of choice. This is a dummy variable that takes value 1 if the price of one of the alternatives in choice set $s$ is given by a discount and 0 otherwise (see the survey form in Table A1 in Appendix A). This is the main variable of interest in our model. If its coefficient turns out to be significantly negative

then we can draw the conclusion that the way prices are specified causes the respondents to have significant difficulties in making the choice.

## 2.3 Estimation and information matrix

The parameters of the model can be estimated by the method of maximum likelihood. The log-likelihood function can be expressed as a constant plus

$$L = \sum_{i=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} y_{ijs} \ln \pi_{ijs}, \tag{2}$$

where $y_{ijs}$ is 1 if respondent $i$ chooses alternative $j$ in choice set $s$ and 0 otherwise, and $\pi_{ijs}$ is the probability that $y_{ijs} = 1$. This probability is given by the formula $\pi_{ijs} = \int_{\mathbb{R}^K} p_{js}(v)\,\phi(v)\,dv$, where

$$p_{ijs}(v) = \frac{\exp\left(\exp\left(c_{is}'\gamma\right) x_{ijs}'\left(\beta + V\sigma\right)\right)}{\sum_{r=1}^{J} \exp\left(\exp\left(c_{is}'\gamma\right) x_{irs}'\left(\beta + V\sigma\right)\right)}$$

and $\phi(v)$ is the probability density function of the $K$-dimensional standard normal distribution. Note that the probability $\pi_{ijs}$ depends on $i$ only through $x_{ijs}$, $j = 1, ..., J$ and $c_{is}$; the latent characteristics $V_{is}$ are integrated out.

In order to design the experiment in an efficient way we need to use the variance of the parameter estimator, or equivalently, the Fisher information matrix (for more details on the efficiency of the experimental design see subsection 3.1). This can be computed as the variance of the first order conditions of the log-likelihood evaluated at the true parameter values. We denote the vector of all parameters by $\theta = (\beta', \sigma', \gamma')'$ and its true

value by $\theta_0 = (\beta_0', \gamma_0', \sigma_0')'$. The information matrix is given by

$$I(X, c; \theta_0) = \sum_{i=1}^{N} \sum_{s=1}^{S} e^{2c_{is}'\gamma_0} \begin{pmatrix} M_{is}'\Pi_{is}^{-1}M_{is} & M_{is}'\Pi_{is}^{-1}Q_{is} & A' \\ Q_{is}'\Pi_{is}^{-1}M_{is} & Q_{is}'\Pi_{is}^{-1}Q_{is} & B' \\ A & B & C \end{pmatrix}, \tag{3}$$

where

$$\begin{aligned} A &= c_{is}\beta_0' M_{is}'\Pi_{is}^{-1}M_{is} + c_{is}\sigma_0' Q_{is}'\Pi_{is}^{-1}M_{is}, \\ B &= c_{is}\beta_0' M_{is}'\Pi_{is}^{-1}Q_{is} + c_{is}\sigma_0' Q_{is}'\Pi_{is}^{-1}Q_{is}, \\ C &= c_{is}\beta_0' M_{is}'\Pi_{is}^{-1}M_{is}\beta_0 c_{is}' + c_{is}\beta_0' M_{is}'\Pi_{is}^{-1}Q_{is}\sigma_0 c_{is}' + c_{is}\sigma_0' Q_{is}'\Pi_{is}^{-1}M_{is}\beta_0 c_{is}' \\ &\quad + c_{is}\sigma_0' Q_{is}'\Pi_{is}^{-1}Q_{is}\sigma_0 c_{is}', \end{aligned}$$

$\Pi_s$ is the diagonal matrix with diagonal $\pi_s = (\pi_{1s}, ..., \pi_{Js})'$,

$$\begin{aligned} M_{is} &= \int_{\mathbb{R}^K} \left[P_{is}(v) - p_{is}(v)\, p_{is}'(v)\right] X_{is}\phi(v)\, dv \quad \text{and} \\ Q_{is} &= \int_{\mathbb{R}^K} \left[P_{is}(v) - p_{is}(v)\, p_{is}'(v)\right] X_{is} V \phi(v)\, dv. \end{aligned}$$

Here $P_{is}(v)$ is the diagonal matrix with diagonal $p_{is}(v) = (p_{i1s}(v), ..., p_{iJs}(v))$ and $X_{is} = (x_{i1s}, ..., x_{iJs})'$. We provide a short derivation of the information matrix in Appendix B.

# 3 Experimental design

We design the experiment with the purpose of collecting data for the model presented above. For this we need to create a so-called design $X$, which is a matrix containing

the characteristics of all hypothetical products from all choice sets given to all the respondents. In order to make the choice tasks easier for the respondents, we adopt a commonly used idea and specify each characteristic by a limited number of levels. Any hypothetical product (i.e., row of $X$) is just a combination of the characteristic levels.

An easy way to construct the design matrix would be to generate it randomly. This strategy is followed in many of the papers mentioned in the Introduction. We note that, although these papers do not explicitly state, the fact that they do not take any design efficiency criterion into account when constructing the design makes their design equivalent to a randomly generated design, at least from a design efficiency point of view. In the next section we show in a Monte Carlo experiment that, assuming that our estimates are sufficiently close to the true values of the parameters, our design is significantly more efficient than a random design.

In order to make the data collection more effective, one respondent is given several choice sets. In our case the number of choice sets we will give to a respondent is $S = 12$. So, strictly speaking, we work with a multi-response design but, in view of the independence of the latent respondent characteristics $V_{is}$ across different choice sets, the design can be regarded as single-response design.

For determining the number of choice sets per respondent and the number of alternatives in a choice set we rely on empirical results by Swait and Adamowicz (2001). Our purpose is to determine the design size so that the complexity of the choice task is moderate. Swait and Adamowicz find that the subjects in the experiment provided responses with no fatigue for the first half of choice sets of a 16-choice set design, in which each choice set has three alternatives. Our interpretation of this finding is that

respondents can evaluate 24 hypothetical products grouped in small choice sets without getting significantly distracted. This is the fact that provided us the reason to give each respondent 12 choice sets with two alternatives.

## 3.1 Design efficiency

The efficiency feature of experimental designs has been the main focus in the statistics literature. A design that is more efficient requires fewer responses and thus reduces the cost of the experiment. The efficiency of a design is measured by the information it provides on the parameters, and this is measured by the Fisher information matrix. The Fisher information matrix provides only an asymptotic approximation of the information but is commonly accepted in the optimal design literature (e.g., Chaloner and Verdinelli 1995). In order to produce a design with improved efficiency, it is desirable to maximize the information matrix. One way to do so is by optimizing a monotonic scalar transformation of its determinant. A design that achieves the optimum is called a D-optimal design.

For our model, and in general for nonlinear models, the determinant of the information matrix depends on the true values of the parameters of the model (see the information matrix in (3)), which are obviously unknown. A natural simple solution is to take some best guess values for the parameters and maximize the design criterion using these. This leads to a so-called locally (D-) optimal design. A better solution is to take a best guess probability distribution of the parameters and maximize the expectation of the D-optimality criterion. This leads to a (D-) optimal Bayesian design.

Taking the above considerations into account we choose the design $X$ and the task

complexity variables $c$ by

$$\min_{X,c} D_B\left(X, c; \theta\right), \text{ where } D_B\left(X, c; \theta\right) = E_\theta\left[\det\left(I\left(X, c; \theta\right)\right)^{-1/d}\right].$$

Here $d$ is the dimension of the information matrix. In this paper we will refer to $D_B\left(\cdot\right)$ as the Bayesian design criterion and to $D_L\left(X, c; \theta\right) = \det\left(I\left(X, c; \theta\right)\right)^{-1/d}$ as the local design criterion. We note that normalizing the determinant by the dimension of the information matrix gives a convenient design criterion because it is inversely proportional to the number of respondents in the experiment. Similar Bayesian D-optimality criteria have been discussed by Atkinson and Donev (1992, Chapter 19), Chaloner and Verdinelli (1995) and Firth and Hinde (1997). The design criterion proposed by the latter authors, though not completely the same as ours, yields the same optimal design as our design criterion.

It is important to observe that there is no need to optimize the design criterion with respect to $c$. As explained at the beginning of this section, $c$ consists of two components. The first component depends on the characteristics (see equation (1)), so its optimality is incorporated in optimizing $D_B$ with respect to $X$. The second component consists of a dummy variable that reflects the price specification. Out of the 12 dummies corresponding to the 12 choice sets given to a respondent, we take 5 values equal to 1 and the rest of 7 equal to 0. We can arbitrarily fix this variable prior to design construction for each 12-tuple of choice sets and determine the optimal design conditional on these. This yields the unconditional optimal design because the information matrix remains unchanged whenever we interchange two different choice

sets. Consequently, we determine the second component of the task complexity variable prior to design construction in an arbitrary way, and keep it fixed. For simplifying notation, we ignore the task complexity variables from the arguments of the design criteria and use $D_L(X;\theta)$ and $D_B(X;\theta)$.

The Bayesian design criterion is a multidimensional integral of the normalized inverse of the determinant of the information matrix, which itself involves multidimensional integrals (see the explanatory formulas below equation (3)). None of these integrals can be computed analytically. The traditional approach to deal with such analytically intractable integrals is to approximate them either by a numerical method (e.g., Gaussian quadratures) or by Monte Carlo simulations. Both methods require a very large number of function evaluations if one wants to achieve reasonably precise approximations. In order to reduce the computational burden we employ a more advanced technique, namely quasi-random simulations (e.g., Niederreiter 1992). Randomized versions of these (Owen 1995) are similar to Monte Carlo simulations in that they use random samples to estimate the integrals, but, instead of using computer-generated pseudo-random samples, they generate the samples so that they fill out the sample space more efficiently. In this paper we employ so-called randomized $(0, m, s)$-nets in bases 4 and 8 (see Niederreiter). These are used to replace pseudo-random samples of the uniform distribution on the unit hypercube and then transformed to produce a sample from any parametric distribution. The asymptotic behavior of these samples for estimating integrals was studied by Tang (1993) and Owen (1994). Their small sample behavior, in the context of a mixed logit model estimation, was investigated by Sándor and Train (2004) and, in the context of estimating multivariate normal probabilities, by Sándor and András (2004). In the next

subsection, when discussing the specific design construction problems, we give more details on the performance of our quasi-random samples as compared to pseudo-random samples.

## 3.2 Design procedure

In order to construct the design we adopt a two-stage procedure. This is motivated by the fact that without having any observations we have little knowledge of the parameter values. Still, we have some expectations about the signs, which we incorporate in their first stage prior distribution. Then, based on the observations collected for the design constructed in the first stage, we obtain more precise information about the parameters, which allows us to construct a more efficient design in the second stage.

In the first stage we construct a Bayesian design $X_1$ based on a uniform prior distribution of the parameters. Then we collect data, say $y_1$, based on this design and estimate the parameters. Denote the estimate by $\widehat{\theta}(X_1, y_1)$, about which it is known that it is asymptotically normally distributed, so we use the asymptotic approximation $\widehat{\theta}(X_1, y_1) \approx N\left(\theta_0, I(X_1; \theta_0)^{-1}\right)$, where we recall that $\theta_0$ is the true value of $\theta$. Then we use this approximate distribution of the estimate to construct a second-stage Bayesian design $X_2$. We do so by determining $X_2$ that solves $\min_{X_2} D_B\left((X_1', X_2')'; \theta\right)$, where $\theta$ has the above distribution. At this stage we replace $\theta_0$ by the estimate $\widehat{\theta}(X_1, y_1)$; this is the way we use the information from the first stage. Once we have $X_2$ we collect data based on it and we estimate the full model. In practice the two-stage procedure in our problem is not as straightforward. We give more details about it in subsection 4.1.

We make a remark about an issue regarding the construction of two-stage opti-

mal Bayesian designs. In order to treat this problem properly one should take a priory into account that the design is constructed in two stages and optimize it accordingly. Zacks (1977) demonstrated such a fully optimal procedure in the case of a one-parameter model. His procedure determines the designs employed in the two stages to be simultanously optimal. For example, when determining the design used in the first stage, the procedure takes optimally into account that there is also a second stage. For our design problem note that $X_2$ determined in the second stage depends on $X_1$, so that we can put $X_2(X_1)$. This implies then the first stage design optimization $\min_{X_1} D_B\left(\left(X_1', X_2(X_1)'\right)'; \theta\right)$, where $\theta$ has a prior distribution. In the case of our problem this minimization is a very complex computational exercise because it is not possible to determine the dependence of $X_2(X_1)$ on $X_1$ analytically. So we do not follow the fully optimal approach, on the one hand because of the above problem, and on the other hand, because the gains in design efficiency may turn out to be minor (see Ridout 1995).

We note that since the main goal of our study is to examine the null hypothesis that price evaluations influence the variance of choice, a procedure more natural than the one we follow would be to construct the design that is optimal for testing this null hypothesis. This would imply a design criterion different from ours, so the optimal design would also be different. Such procedures have been discussed in the literature (Ponce De Leon and Atkinson 1991). The reason that we do not follow such a procedure is that our goal is to obtain estimates for all parameters of the model in order to judge more globally the outcome of our experiment. Our goal is also to assess the correctness of the model for the experimental data generating process by verifying the estimates,

for example, by their signs.

## 3.3 The algorithm

The algorithm we employ for optimizing the design criterion has been used successfully to construct designs for choice experiments. It is related to the exchange algorithm proposed by Fedorov (1972) and modified by Cook and Nachtsheim (1980) (see Atkinson and Donev 1992, Chapter 15, for a comprehensive review of algorithms for design construction). For choice experiments this exchange algorithm has been tailored according to the specific features of the discrete choice models used.

Here we describe the main features of the algorithm. As it will become clear below, the exchange idea is adapted by modifying the choice sets of a randomly generated starting design successively. There are two procedures that are employed for this purpose, that is, swapping and cycling. Swapping (Huber and Zwerina 1996) works by evaluating the criterion function after each successive swap of the characteristics levels within the choice sets and if no improvement is found then it goes to the next choice set, and so on. Cycling (Sándor and Wedel 2001, 2002) modifies the levels of a characteristic in a choice set successively so that eventually all level-pairs of this characteristics are considered, and evaluates the criterion function after each modification. After considering the first characteristic in the first choice set it continues with the first characteristic in the next choice sets; after exhausting all choice sets it goes to the other characteristics. For both swapping and cycling, if an improvement is found, the procedure goes back to the first characteristic of the first choice set. The procedures stop when they arrive at the last characteristic of the last choice set and find no improvement in the criterion function.

We note that swapping is more restrictive than cycling because it keeps the same levels of a characteristic in a choice set, namely, it only changes them across the alternatives. In the algorithm we apply first swapping and then cycling. This way we try to avoid obtaining a design that is a local optimum of the design criterion, but still there is no guarantee that the globally optimal design is obtained.

Due to this fact, this algorithm, as most exchange algorithms, does not yield an optimal design in general but rather a design with improved efficiency (see also Atkinson and Donev 1992, Chapter 15). Its efficiency can be assessed to some extent by applying it in special cases. Sándor and Wedel (2001, footnote 2) consider a class of very small designs for which all different designs could be evaluated. They find that the algorithm yields an optimal design or a very closely optimal design. Sándor and Wedel (2003) verify the efficiency of the algorithm in a design class for paired comparison models, which are a restricted version of mixed logit. Street et al. (2001) and Street and Burgess (2004) derived the optimal designs, which can be compared to the outcome of the swapping part of the algorithm. Just like above, this turns out to produce either an optimal design or a very closely optimal design.

Applying swapping and cycling to all choice sets of our design is not a feasible task with the currently available computing power. Therefore, we reduce the number (i.e., $N \cdot 12$) of different choice sets to 120. There are reasons to believe that the efficiency per one respondent remains approximately the same (Sándor and Wedel 2003). Further, we employ a 'greedy' approach by splitting the 120-choice set design into 10 smaller subdesigns of 12 choice sets each and construct the design in a sequential fashion. More precisely, we apply swapping and cycling to a randomly generated starting design $X_0$

with 12 choice sets. This way we obtain the first subdesign $X_1^*$. Then we consider the design with two subdesigns $X_1^*$ and $X_0$ and we apply swapping and cycling to $X_0$ while keeping $X_1^*$ fixed. This way we obtain the second subdesign $X_2^*$. Then we continue in this fashion until the 10'th subdesign $X_{10}^*$ is constructed. The computational gain with respect to the case of full 120 choice sets is that at each stage we apply swapping and cycling only to 12 choice sets. We note that this sequential feature of the algorithm provides yet another reason why our design cannot be guaranteed to be globally optimal.

# 4 Data collection and results

We collect data from students on mobile phone preferences. For the hypothetical mobile phones we use the following characteristics: price, price per minute, extras, network, SMS price and design. The levels of these characteristics are presented in Table A2 in Appendix A. The first three characteristics have three levels and the last three characteristics have two levels. The characteristics price and price per minute are assumed to be quantitative, while the characteristic "extras" is assumed to be qualitative (see Table A2 for its coding).[3]

Because the student population, which is considered for data collection, is expected to be less heterogeneous than the whole population of consumers, and also for reasons of parsimony in the design construction and estimation, we assume that preferences are heterogeneous only with respect to certain characteristics (price, extras1 and design).

[3]We note that the cycling procedure tends to select only the extreme levels in the case of a three-level quantitative characteristic. For reasons of realism our purpose is to maintain all three levels of the characteristics price and price per minute. Therefore we modified slightly the algorithm described above by omitting cycling to these two characteristics. This modification reduces the efficiency of the design we obtained but enhances the realism of our experiment.

This way we obtained a model with 12 parameters of which two are parameters related to choice task complexity and three are standard deviation parameters.

## 4.1 Specific issues

As explained above, we construct our experimental design in two stages. Here we discuss specific issues regarding the first and second stage designs. We pay special attention to a problem we believe is common for design construction for the mixed logit or probit models. This is a problem that may occur due to the fact that the estimates obtained in the first stage are based on a too small sample and hence they may not even exist.

In the first stage we construct the design by assuming uniform prior distributions for the mean parameters (presented in Table A2) and values equal to 1 for the three standard deviation parameters. We optimize the Bayesian design criterion $D_B$ with respect to the levels of the characteristics. So the design problem in the first stage involves a mixture of Bayesian and local optimality, or in other words, Bayesian optimality with degenerate prior distributions for the standard deviation parameters. We do not consider optimizing the design with respect to the two task complexity variables in the first stage.

Based on the design constructed in the first stage we collected 116 observations. For numerical maximization of the log-likelihood we employed the nonderivative simplex search algorithm.[4] In several runs of the maximization algorithm with random starting values this algorithm found several local maxima of the log-likelihood. Out of these there is a sequence of local maximum points which go away from the origin while their

[4]The Gauss code for this algorithm written by Bo Honore and Ekaterini Kyriazidou is publicly available on the internet.

function values slightly increase (see Table A3 in Appendix A). Hence, we are tempted to draw the conclusion that there exists no global maximum to the log-likelihood, or in other words, the maximum likelihood estimator does not exist. We note that this is not an uncommon phenomenon when estimating mixed logit or probit models with a small sample size, so researchers who intend to construct designs with improved efficiency for these models may encounter the same problem.

Still it is possible to use the information from the first stage observations, and for this we propose the following procedure. First we note that if we knew the standard deviation parameters then we could estimate the rest of the parameters and the (asymptotic) distribution of these estimates using the restricted mixed logit model with task complexity variables. This restricted model turns out to have excellent convergence properties. If we were to follow this procedure we could assume some (second stage) prior values for the standard deviation parameters and use the corresponding (asymptotic) distribution of the obtained estimates of the other parameters as the (second stage) prior distribution. This would then imply a mixture of Bayesian and local optimality of the design, similar to the first stage. Formally, we would optimize the criterion

$$E_{\beta,\gamma}\left[D_L\left(\left(X_1', X_2'\right)'; \beta\left(\sigma\right), \gamma\left(\sigma\right), \sigma\right)\right]$$

with respect to $X_2$. Based on the asymptotic normality of the first stage estimates, this can be approximated by

$$E_u\left[D_L\left(\left(X_1', X_2'\right)'; \mu\left(\sigma\right) + \Lambda\left(\sigma\right) u, \sigma\right)\right],$$

where $\mu(\sigma)$ and $\Lambda(\sigma)$ are the mean and the Cholesky decomposition of the asymptotic variance of the estimates of $\beta_0$ and $\gamma_0$ for a given value of $\sigma$, and $u$ is a standard normal random vector. However, we would like to use a fully Bayesian design criterion. Therefore, instead of assuming some value for the standard deviation parameters, we assume a distribution for them. Formally, the design criterion in this case becomes

$$E_{u,\sigma}\left[D_L\left(\left(X_1', X_2'\right)'; \mu(\sigma) + \Lambda(\sigma) u, \sigma\right)\right]. \tag{4}$$

In practice this expectation needs to be approximated. We note that for the integrals involved in the information matrix (below equation (3)) we employed a randomized $(0, 3, s)$-net in base 4 of dimension $s = 3$. This sample has been shown to reduce computation time by a factor of eight with respect to pseudo-random samples in a mixed logit estimation problem (Sándor and Train 2004). Abstracting from these integrals, the expectation in (4) involves a 12-dimensional integral. We use two independent samples of size 64 each for simulating $u$ and $\sigma$, namely, a randomized $(0, 2, s)$-net in base 8 of dimension $s = 9$ for $u$ and a randomized $(0, 3, s)$-net in base 4 of dimension $s = 3$ for $\sigma$. We assume independent normal distributions with means 1.5 and standard deviations 0.7 for its components. We implement the application of these samples by drawing 64 values $\sigma_1, ..., \sigma_{64}$ for $\sigma$, estimating the reduced model and computing $\mu(\sigma_i)$ and $\Lambda(\sigma_i)$ for each $i = 1, ..., 64$, then drawing 64 values $u_1, ..., u_{64}$ for $u$, and finally computing $\frac{1}{64}\sum_{i=1}^{64} D_L\left(\left(X_1', X_2'\right)'; \mu(\sigma_i) + \Lambda(\sigma_i) u_i, \sigma_i\right)$.

In a small simulation experiment we compare the performance of these samples to pseudo-random samples of the same size. We use a specific design and prior values

obtained from the first stage as described above. We obtain estimates of the mean of the Bayesian design criterion 0.0684 and 0.0656 with standard deviations 0.00062 and 0.00226 with the quasi-random samples and the pseudo-random samples, respectively. This suggests that we need pseudo-random samples of size about 840 in order to obtain the same precision as with quasi-random samples. In other words, using pseudo-random samples we need about 13 times as much computing time as with our quasi-random samples. This is a remarkable reduction of computing time, especially because constructing the second stage design takes about 55 hours (on a Pentium II PC). We believe that quasi-random samples yield similar computational gains in other Bayesian design construction problems as well.

For each $\sigma_i$ for $i = 1, ..., 64$ we estimate the other parameters of the model. Based on several maximization runs with random starting values we find that there is a unique maximum point of the reduced log-likelihood in each case. This property appears to hold in spite of the fact that, due to the inclusion of the task complexity variables, the concavity of the log-likelihood cannot be proved theoretically in the same way as for the restricted mixed logit.

## 4.2 Results

Based on the design constructed in the second stage we collect 350 observations. Then we estimate the model using the data collected in both stages. In the estimation, in order to achieve very high precision for the integral estimates of $\pi_{ijs}$ involved in the log-likelihood (2), we employed quasi-random samples based on a $(0, 4, s)$-net of dimension $s = 3$ and sample size 256. The estimates and their standard errors are presented

in Table 1 under the column '2nd stage.' The parameter estimates in this table, as compared to their standard errors, suggest that all variables are relevant in explaining the mobile phone preferences of the students, with the exception of task complexity and SD design. The parameter estimates have the expected signs, in the sense that lower prices are preferred, as well as that one seems to prefer more extras, a KPN or Vodafone network and a trendy design of the phone. As outlined in section 2.2, our focal variable is the "price specification", and we expect that, due to potentially distracting information, its parameter shall be negative, meaning that respondents find it difficult to evaluate, for example, 135 euro versus 150 euro minus 10 per cent discount. Table 1 shows that this parameter is estimated as –0.191, with an associated standard error of 0.093. Hence, we obtain empirical evidence that price evaluations can be difficult indeed. We do not obtain empirical evidence, however, that the choice task, as we measure it, is difficult to the respondents. We believe that this is due to the fact that the number of alternatives in a choice set is only two.

We conduct a simulation experiment in which we compare the local design criterion values $D_L$ and standard errors of the two-stage design, a randomly generated design and the first stage design. We recall that a randomly generated design can be regarded as equivalent to designs constructed without statistical efficiency considerations. Most experimental design studies on demand use such designs. This simulation experiment is motivated by the question whether it is worthwhile to construct designs based on efficiency considerations.

Table 1. Estimates and standard errors for the two-stage, randomly generated and 1st stage designs

| | 2nd stage | | Random | 1st stage |
|---|---|---|---|---|
| | Estimate | St. error | St. error | St. error |
| Price | -1.133 | 0.314 | 0.649 | 0.376 |
| Price/min | -0.922 | 0.262 | 0.531 | 0.321 |
| Extras1 | -0.414 | 0.118 | 0.233 | 0.147 |
| Extras2 | -0.138 | 0.052 | 0.087 | 0.074 |
| Network | -0.235 | 0.067 | 0.136 | 0.080 |
| SMS price | -0.239 | 0.068 | 0.139 | 0.080 |
| Design | 0.211 | 0.064 | 0.122 | 0.089 |
| Task complexity | 0.098 | 0.255 | 0.787 | 0.351 |
| Price specification | -0.191 | 0.093 | 0.241 | 0.090 |
| S.D. Price | 0.868 | 0.296 | 0.563 | 0.327 |
| S.D. Extras1 | 0.642 | 0.266 | 0.431 | 0.295 |
| S.D. Design | 0.459 | 0.290 | 1.199 | 0.349 |
| Local design criterion $D_L$ | | 0.0037 | 0.0065 | 0.0046 |

The results are presented in Table 1. In evaluating the local design criteria and the standard errors we assume that the estimates are sufficiently close to the true values of the parameters. We generated randomly a design with as many subdesigns with 12 choice sets as the number of respondents we have altogether, that is, 466. We replicate this 100 times and compute the mean of the standard errors and of the local design criterion values. As expected, the two-stage design has the lowest value of the local design criterion values; it is about 1.76 times lower than that of the randomly generated design. This implies that, in order to achieve the efficiency (as measured by the local design criterion) of the two-stage design, we need about 819 respondents instead of 466 if we use the randomly generated design. The local design criterion of the two-stage design is about 1.24 times lower than that of the first stage design. This means that, in order to achieve the efficiency of the two-stage design, we need about 579 respondents

instead of 466 if we use the first stage design.

Regarding the standard errors we can draw a similar conclusion. The standard errors corresponding to the two-stage design are systematically lower than the standard errors corresponding to the first stage design and these are further lower than the mean of the standard errors corresponding to the randomly generated design.

# 5 Conclusion

In our experiment we have obtained an affirmative answer to the question whether price specification affects consumer choice. An important issue is whether we can draw the conclusion that this is so in reality. The results regarding the other parameters of the model make us believe that the model is a reasonable approximation of the respondents' preferences. So we believe that our conclusion on price specification is true in reality as well.

The simulation results that compare our design to a randomly generated design show that it is indeed worthwhile to construct two-stage designs because the gains in estimation precision are significant. These results serve as an argument for the claim that it is worthwhile to construct designs based on efficiency considerations for experiments used for discrete choice models. So in this respect discrete choice models are not different from the many statistical models considered for optimal design in the statistics literature.

A challenging topic for future research is to find out how close to the optimal design are the designs obtained by our procedure. A possible solution to this problem is offered by applying Wynn's (1970) sequential design construction procedure because this yields

designs that converge to the optimal design. The main challenge here is to implement the procedure in a computationally parsimonious way, given the computational burden implied by the mixed logit or probit model.

Another interesting topic is the design of experiment for models that involve demographic characteristics of consumers (as, for example, in van Ophem et al. 1999). This generates different types of problems depending on whether the researcher is free to determine an optimal sample of respondents or not. The former case implies an optimal design problem similar to the problem of optimal Monte Carlo simulations. In the latter case the researcher may or may not know the sample prior to the experiment. The case when he/she does not is the more challenging one.

An interesting optimal design problem occurs when one intends to estimate consumer preferences from a combination of real-life and experimental data (e.g., Adamowicz et al. 1997). Real-life data provide estimates whith a distribution that can naturally be employed as the prior distribution for constructing the Bayesian design. Hence in this situation the design procedure may reduce to a single stage if the prior distribution is sufficiently informative for constructing a good design.

# References


Adamowicz, W., J. Swait, P. Boxall, J. Louviere and M. Williams (1997), 'Perceptions versus objective measures of envirobmental quality in combined revealed and stated preference models of environmental valuation', *Journal of Environmental Economics and Management* **32**, 65–84.

Allenby, G.M. and P.J. Lenk (1994), 'Modeling household purchase behavior with logistic normal regression', *Journal of the American Statistical Association* **89**, 1218–31.

Arora, N. and J. Huber (2001), 'Improving parameter estimates and model prediction by aggregate customization in choice experiments', *Journal of Consumer Research* **28**, 273–83.

Atkinson, A.C. and A.N. Donev (1992), *Optimum Experimental Designs*, London: Oxford.

Berry, S., J. Levinsohn and A. Pakes (1995), 'Automobile prices in market equilibrium', *Econometrica* **63**, 841–890.

Bettman, J.R., E.J. Johnson, M.F. Luce and J.W. Payne (1993), 'Correlation, conflict and choice', *Journal of Experimental Psychology: Learning, Memory and Cognition* **19**, 931–951.

Brownstone, D. and K. Train (1999), 'Forecasting new product penetration with flexible substitution patterns', *Journal of Econometrics* **89**, 109–29.

Chaloner, K. and I. Verdinelli (1995), 'Bayesian experimental design: A review', *Statistical Science* **10 (3)**, 273–304.

Cook, R.D. and C.J. Nachtsheim (1980), 'A comparison of algorithms for constructing exact D-optimum designs', *Technometrics* **22**, 315–24.

DeShazo, J.R. and G. Fermo (2002), 'Designing choice sets for stated preference methods: The effects of complexity on choice consistency', *Journal of Environmental Economics and Management* **44**, 123–43.

Fedorov, V.V. (1972), *Theory of Optimal Experiments*, Academic Press: New York.

Firth, D. and J.P. Hinde (1997), 'On Bayesian D-optimum design criteria and the equivalence theorem in non-linear models', *Journal of the Royal Statistical Society B* **59**, 793–97.

Hensher, D.A. and Ch. Sullivan (2003), 'Willingness to pay for road curviness and road type', *Transportation Research D* **8**, 139–55.

Huber, J. and K. Zwerina (1996), 'The importance of utility balance in efficient choice design', *Journal of Marketing Research* **33**, 307–17.

Johnson, E.J. and J.W. Payne (1985), 'Effort and accuracy in choice', *Management Science* **31**, 395–414.

Layton, D.F. and G. Brown (2000), 'Heterogenous preferences regarding global climate change', *Review of Economics and Statistics* **82**, 616–24.

Louviere, J.J. (2003), Complex statistical choice models: Are the assumptions true, and if not, what are the consequences? Presented at Discrete Choice Workshop in Health Economics, University of Oxford.

Louviere, J.J. and G. Woodworth (1983), 'Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data', *Journal of Marketing Research* **20**, 350–67.

McFadden, D. and K. Train (2000), 'Mixed MNL models for discrete response', *Journal of Applied Econometrics* **15**, 447–70.

Montepiedra, G. and A.B. Yeh (1998), 'A two-stage strategy for the construction of D-optimal experimental designs', *Communications in Statistics–Simulation and Computation* **27**, 377–401.

Niederreiter, H. (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, Philadelphia: SIAM.

Owen, A.B. (1994), 'Lattice sampling revisited: Monte Carlo variance of means over randomized orthogonal arrays', *Annals of Statistics* **22**, 930–945.

Owen, A.B. (1995), Randomly permuted (t,m,s)-nets and (t,s)-sequences, *in* H. Niederreiter and P.J.-S. Shiue, ed., 'Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing', New York: Springer-Verlag, pp. 299–317.

Ponce De Leon, A.C. and A.C. Atkinson (1991), 'Optimum experimental design for discriminating between two rival models in the presence of prior information', *Biometrika* **78**, 601–608.

Revelt, D. and K. Train (1998), 'Mixed logit with repeated choices: Households' choices of appliance efficiency level', *Review of Economics and Statistics* **80**, 647–57.

Ridout, M. (1995), 'Three stage designs for seed testing experiments', *Journal of the Royal Statistical Society C* **44**, 153–62.

Rubinstein, A. (1998), *Modeling Bounded Rationality*, Cambridge MA: MIT Press.

San-Miguel, F., M. Ryan and A. Scott (2002), 'Are preferences stable? The case of health care', *Journal of Economic Behavior and Organization* **48**, 1–14.

Sándor, Z. and K. Train (2004), 'Quasi-random simulation of discrete choice models', *Transportation Research B* . To appear.

Sándor, Z. and M. Wedel (2001), 'Designing conjoint choice experiments using managers' prior beliefs', *Journal of Marketing Research* **38**, 430–44.

Sándor, Z. and M. Wedel (2002), 'Profile construction in experimental choice designs for mixed logit models', *Marketing Science* **21**, 55–75.

Sándor, Z. and M. Wedel (2003), Differentiated Bayesian conjoint choice designs, Technical Report ERS-2003-016-MKT, ERIM.

Sándor, Z. and P. András (2004), 'Alternative sampling methods for multivariate normal probabilities', *Journal of Econometrics* . To appear.

Scott, A. (2001), 'Eliciting GP's preferences for pecuniary and non-pecuniary job characteristics', *Journal of Health Economics* **20**, 329–47.

Street, D.J., D.S. Bunch and B.J. Moore (2001), 'Optimal designs for $2^k$ paired comparison experiments', *Communications in Statistics–Theory and Methods* **30**, 2149–71.

Street, D.J. and L. Burgess (2004), 'Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments', *Journal of Statistical Planning and Inference* **118**, 185–199.

Swait, J. and W. Adamowicz (2001), 'The influence of task complexity on consumer choice: A latent class model of decision strategy switching', *Journal of Consumer Research* **28**, 135–48.

Tang, B. (1993), 'Orthogonal array-based Latin hypercubes', *Journal of the American Statistical Association* **88**, 1392–1397.

Tyebjee, T. (1979), 'Response time, conflict and involvement in brand choice', *Journal of Consumer Research* **6**, 295–304.

van Ophem, H., P. Stam and B. van Praag (1999), 'Multichoice logit: Modeling incomplete preference rankings of classical concerts', *Journal of Business and Economic Statistics* **17**, 117–28.

Wynn, H.P. (1970), 'The sequential generation of D-optimum experimental designs', *The Annals of Mathematical Statistics* **41**, 1655–64.

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

# Appendix A: Additional information

Table A1. Example of a design with six choice sets

| | Product A | Product B |
|---|---|---|
| Price | € 135 | € 189, now with 10% discount |
| Network | other than KPN and Vodafone | other than KPN and Vodafone |
| Price/minute | € 0.25 | € 0.35 |
| SMS price | € 0.17 | € 0.23 |
| Design | trendy | trendy |
| Extras | games & internet & camera | games & internet & camera |
| Price | € 100 | € 170 |
| Network | KPN or Vodafone | other than KPN and Vodafone |
| Price/minute | € 0.25 | € 0.30 |
| SMS price | € 0.17 | € 0.23 |
| Design | trendy | basic |
| Extras | games & internet | games & internet & camera |
| Price | € 135 | € 125, now with 20% discount |
| Network | KPN or Vodafone | other than KPN and Vodafone |
| Price/minute | € 0.25 | € 0.35 |
| SMS price | € 0.17 | € 0.17 |
| Design | basic | trendy |
| Extras | games | games & internet & camera |
| Price | € 100 | € 170 |
| Network | other than KPN and Vodafone | KPN or Vodafone |
| Price/minute | € 0.30 | € 0.25 |
| SMS price | € 0.23 | € 0.17 |
| Design | basic | trendy |
| Extras | games & internet | games |
| Price | € 135 | € 170 |
| Network | other than KPN and Vodafone | other than KPN and Vodafone |
| Price/minute | € 0.30 | € 0.35 |
| SMS price | € 0.17 | € 0.23 |
| Design | trendy | trendy |
| Extras | games & internet & camera | games |
| Price | € 170 | € 150, now with 10% discount |
| Network | KPN or Vodafone | other than KPN and Vodafone |
| Price/minute | € 0.25 | € 0.30 |
| SMS price | € 0.23 | € 0.17 |
| Design | basic | trendy |
| Extras | games & internet | games & internet & camera |

Table A2. Characteristics levels, their coding and 1st stage priors

| Variable | Presented | Coded | | 1st stage prior | |
|---|---|---|---|---|---|
| Price (€) | 100 | -1 | | [-1,0] | |
| | 135 | 0 | | | |
| | 170 | 1 | | | |
| Price/min (€) | 0.25 | -1 | | [-2,0] | |
| | 0.30 | 0 | | | |
| | 0.35 | 1 | | | |
| Extras | games | 1 | 0 | [-4,0] | [-2,0] |
| | games & internet | 0 | 1 | | |
| | games & internet & camera | -1 | -1 | | |
| Network | KPN or Vodafone | -1 | | [-1,1] | |
| | other than KPN and Vodafone | 1 | | | |
| SMS price (€) | 0.17 | -1 | | [-1,0] | |
| | 0.23 | 1 | | | |
| Design | basic | -1 | | [0,2] | |
| | trendy | 1 | | | |

Table A3. Local minima of the log-likelihood in the first stage

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Price | -3.539 | -4.570 | -8.216 | -14.340 | -20.058 | -37.789 | -55.188 |
| Price/min | -4.384 | -5.732 | -10.917 | -18.308 | -26.366 | -52.278 | -76.725 |
| Extras1 | -1.698 | -2.084 | -4.234 | -7.082 | -9.946 | -20.721 | -30.133 |
| Extras2 | -0.467 | -0.556 | -0.725 | -2.098 | -2.326 | -4.543 | -6.485 |
| Network | -0.986 | -1.296 | -2.499 | -4.253 | -6.686 | -12.852 | -18.807 |
| SMS price | -0.583 | -0.767 | -1.598 | -2.277 | -3.489 | -6.805 | -9.731 |
| Design | 0.749 | 0.913 | 1.523 | 3.439 | 4.139 | 8.099 | 12.039 |
| Task complexity | -1.235 | -1.443 | -1.861 | -2.543 | -2.814 | -3.472 | -3.831 |
| Price specification | -0.223 | -0.265 | -0.438 | -0.372 | -0.438 | -0.494 | -0.523 |
| S.D. Price | 2.499 | 3.654 | 8.109 | 11.891 | 18.370 | 35.978 | 54.236 |
| S.D. Extras1 | -1.292 | -1.436 | -4.067 | -7.759 | -10.834 | -22.614 | -32.647 |
| S.D. Design | 3.056 | 3.973 | 7.121 | 13.519 | 18.793 | 36.076 | 52.526 |
| Log-lokelihood | -43.948 | -43.930 | -43.913 | -43.872 | -43.866 | -43.843 | -43.840 |

# Appendix B: Derivation of the information matrix

In this appendix we derive the Fisher information matrix for the mixed logit model with task complexity given in (3). For saving notation, we derive the formulas only for one respondent and one choice set; then the information matrix in (3) will be the sum over all consumers and choice sets. Also, we write integrals like $\int_{\mathbb{R}^K}(\cdot)\phi(v_1)...\ \phi(v_K)dv$ as $\int(\cdot)d\Phi$ (e.g., $\int_{\mathbb{R}^K} p_j(v)\,\phi(v_1)...\phi(v_K)dv \equiv \int p_j d\Phi$).

The log-likelihood is a constant plus

$$L = \sum_{j=1}^{J} y_j \cdot \ln \pi_j = y' \ln \pi,$$

where $y_j$ is 1 if the consumer chooses $j$ and 0 otherwise, $\pi_j$ is the probability that $y_j = 1$, $y = (y_1, ..., y_J)'$ and $\pi = (\pi_1, ..., \pi_J)'$. The Fisher information matrix is given by the formula $I(X, c; \theta_0) = E\left[\frac{\partial L}{\partial \theta}\frac{\partial L}{\partial \theta'}\right]$, where $X$ and $c$ is the design matrix and the the task complexity vector corresponding to the choice set. Using the components $\beta$, $\sigma$, $\gamma$ of $\theta$, we can write the information matrix as

$$I(X, c; \theta_0) = \begin{pmatrix} E\left[\frac{\partial L}{\partial \beta}\frac{\partial L}{\partial \beta'}\right] & & \\ E\left[\frac{\partial L}{\partial \sigma}\frac{\partial L}{\partial \beta'}\right] & E\left[\frac{\partial L}{\partial \sigma}\frac{\partial L}{\partial \sigma'}\right] & \\ E\left[\frac{\partial L}{\partial \gamma}\frac{\partial L}{\partial \beta'}\right] & E\left[\frac{\partial L}{\partial \gamma}\frac{\partial L}{\partial \sigma'}\right] & E\left[\frac{\partial L}{\partial \gamma}\frac{\partial L}{\partial \gamma'}\right] \end{pmatrix}.$$

The upper triangular part is determined by the lower triangular part by the symmetry of the information matrix. We need to compute the first order derivatives of $L$.

We note that for a parameter $\lambda$ vector that is one of $\beta$, $\sigma$, $\gamma$

$$\frac{\partial L}{\partial \lambda} = \left( y' \frac{\partial \ln \pi}{\partial \lambda'} \right)' = \left( \frac{\partial \pi}{\partial \lambda'} \right)' \Pi^{-1} y, \tag{5}$$

where $\Pi$ is the diagonal matrix with diagonal $\pi$. So we need to compute $\frac{\partial \pi}{\partial \lambda'}$ for $\lambda = \beta$, $\sigma$, $\gamma$. Since $\pi = \int p d\Phi$, we have that $\frac{\partial \pi}{\partial \lambda'} = \int \frac{\partial p}{\partial \lambda'} d\Phi$. The vector $p$ is defined as $p = (p_1, ..., p_J)'$ with components

$$p_j(v) = \frac{\exp\left(\exp(c'\gamma) x_j' (\beta + V\sigma)\right)}{\sum_{r=1}^{J} \exp\left(\exp(c'\gamma) x_r' (\beta + V\sigma)\right)}.$$

Hence we obtain

$$\begin{aligned}
\frac{\partial p}{\partial \beta'} &= e^{c'\gamma} (\Pi - \pi\pi') X, \\
\frac{\partial p}{\partial \sigma'} &= e^{c'\gamma} (\Pi - \pi\pi') XV, \\
\frac{\partial p}{\partial \gamma'} &= e^{c'\gamma} (\Pi - \pi\pi') X (\beta + V\sigma) c'.
\end{aligned}$$

Using these, the formula $\frac{\partial \pi}{\partial \lambda'} = \int \frac{\partial p}{\partial \lambda'} d\Phi$ and (5) we obtain

$$\begin{aligned}
\frac{\partial L}{\partial \beta} &= e^{c'\gamma} \int X' (\Pi - \pi\pi') d\Phi \Pi^{-1} y, \\
\frac{\partial L}{\partial \sigma} &= e^{c'\gamma} \int VX' (\Pi - \pi\pi') d\Phi \Pi^{-1} y, \\
\frac{\partial L}{\partial \gamma} &= e^{c'\gamma} \int c (\beta + V\sigma)' X' (\Pi - \pi\pi') d\Phi \Pi^{-1} y.
\end{aligned}$$

Now we are able to compute the components of the information matrix. For this we

introduce the notation

$$M = \int (\Pi - \pi\pi') \, X d\Phi \quad \text{and} \quad Q = \int (\Pi - \pi\pi') \, XV d\Phi$$

and use the fact that $E[yy'] = \Pi$. Then

$$E\left[\frac{\partial L}{\partial \beta}\frac{\partial L}{\partial \beta'}\right] = e^{2c'\gamma} M'\Pi^{-1} E[yy'] \Pi^{-1} M = e^{2c'\gamma} M'\Pi^{-1} M.$$

The other components can be computed in a similar way. Then the information matrix becomes

$$I(X, c; \theta_0) = e^{2c'\gamma_0} \begin{pmatrix} M'\Pi^{-1}M & & \\ Q'\Pi^{-1}M & Q'\Pi^{-1}Q & \\ A & B & C \end{pmatrix},$$

where

$$\begin{aligned} A &= c\beta_0' M'\Pi^{-1}M + c\sigma_0' Q'\Pi^{-1}M, \\ B &= c\beta_0' M'\Pi^{-1}Q + c\sigma_0' Q'\Pi^{-1}Q, \\ C &= c\beta_0' M'\Pi^{-1}M\beta_0 c' + c\beta_0' M'\Pi^{-1}Q\sigma_0 c' + c\sigma_0' Q'\Pi^{-1}M\beta_0 c' \\ &\quad + c\sigma_0' Q'\Pi^{-1}Q\sigma_0 c'. \end{aligned}$$