Stockholms universitet

Research Report

Department of Statistics

No. 2011:5

Minimax Optimality and the Trinomial Spike Model

Linda Wänström
Ellinor Fackle-Fornius

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Minimax Optimality and the Trinomial Spike Model

Linda Wänström and Ellinor Fackle-Fornius

Abstract

This paper investigates designs of contingent valuation experiments in which the researcher is interested in knowing whether respondents have positive willingness to pay and if so, if they are willing to pay a certain amount for a specified good. A trinomial spike model is used to model the response. Optimal design for this model depends on values of unkown parameters, which means that it cannot be used in practice. A purpse of this paper is to show how information about the parameters, e.g. from pilot studies, can be used to find a design. Specifically, a minimax approach is proposed, in which the best guaranteed value of the criterion function is sought under the assumption that the parameter values are within certain regions. The proposed methodology is illustrated in an application where the value of environmentally friendly produced clothes is evaluated.

1 Introduction

Contingent valuation experiments (CVE) are commonly used to estimate the value of non-market goods or services, such as environmental resources. The goal is to assess willingness to pay (WTP) in the population. In a CVE respondents are presented with a hypothetical scenario under which the good or service is to be offered and then exposed to a bidding scenario. Several kinds of bidding scenarios exist but most widely used is the dichotomous choice experiment. The setup may be a single-bound approach, where one bid is presented which the respondent may accept or reject, or a multiple-bound approach, where subsequent follow-up bids are given dependent on the previous answer. See for example Hanemann (1984) and Hanemann and Kanninen (1996) for details. Models that are often used for these setups include standard models for binary data, e.g. the logistic model. Such models describe how the probability that a bid is accepted depends on the bid size and possibly some other relevant variables. Often, rejection of negative bids is implausible, however a large proportion of individuals may have zero WTP. This can be modelled by a spike model which is basically a truncated version of the logistic model with support only for positive bid values, with a "spike" at zero that models the probability that an individual has zero WTP. Different spike models are described in Kriström (1997). If we want to incorporate the possibility that an individual has zero WTP by explicitly asking such a question, a suitable model is the trinomial spike model.

The design issues, when planning a CVE, involve selecting the number of bids, the sizes of the bids, as well as the number of respondents allocated to each bid, in order to maximize precision in the estimation. The theory of optimal design provides methods to plan an experiment to achieve optimality with respect to different inferential objectives, see e.g. Atkinson et al. (2007) or Silvey (1980). For example, if the objective is to estimate the model parameters with high precision it is reasonable to minimize the generalized variance of the parameter estimators, yielding a design that is D-optimal. As another example, if the objective is to estimate some function of the model parameters, e.g. median WTP, minimizing the variance of the associated estimator yields a design that is c-optimal. A problem here is that optimal designs for models for binary data generally require knowledge of the true model parameters. This means that it cannot be used in practice. A purpse of this paper is to show how preliminary information about the parameters, e.g. from pilot studies, can be used to find a design. We propose the minimax approach and demonstrate how such designs are derived for the trinomial spike model. A CVE to estimate WTP for environment friendly production of clothes will serve as an illustration of the methods and results. In addition, we explore some of the patterns in the designs for the trinomial spike model given different prior information about the true parameters.

2 The trinomial spike model

In some situations we are particularly interested in the proportion of individuals with zero WTP in addition to other, positive, levels of WTP. In these situations we can use a particular setup in which two questions are asked to each respondent: if the respondent accepts *any* positive cost and if the respondent accepts or rejects the bid A (see Kriström, Nyquist and Haneman (1992) and Johansson, Kriström and Nyquist (1993) for a description of this setup). The response variable $Y = \begin{pmatrix} Y_0, & Y_1, & Y_2 \end{pmatrix}$ has a trinomial distribution with $E(Y) = \begin{pmatrix} \pi_0, & \pi_1(A), & \pi_2(A) \end{pmatrix}^T$, where

$$\begin{aligned}
Y_0 &= \begin{cases} 1 & \text{if the respondent rejects any positive cost} \\ 0 & \text{otherwise} \end{cases} \\
Y_1 &= \begin{cases} 1 & \text{if the respondent is willing to pay a positive amount but rejects the bid } A \\ 0 & \text{otherwise} \end{cases} \\
Y_2 &= \begin{cases} 1 & \text{if the respondent accepts the bid } A \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

The probability that a randomly selected individual accepts a bid A, $\pi_2(A)$, is a function of the bid size A. The functional form of $\pi_2(A)$ describes the relationship between the response probability and the bid size. It is here defined as a logistic function

$$\pi_2(A) = P(Y_2 = 1) = P(WTP \geq A) = \frac{e^{\beta_0 + \beta_1 A}}{1 + e^{\beta_0 + \beta_1 A}}, \qquad A > 0. \tag{1}$$

The probability that a randomly selected individual has zero WTP, π_0, is defined as

$$\pi_0 = P(Y_0 = 1) = P(WTP = 0) = (1 + e^{\beta_0})^{-1}. \tag{2}$$

Note that this probability is not related to the bid size. The probability that a randomly selected individual is willing to pay som positive amount less than A, $\pi_1(A) = P(Y_1 = 1) = P(0 < WTP < A)$, follows from 1 and 2 since $\pi_0 + \pi_1(A) + \pi_2(A) = 1$.

The trinomial spike model is a logit model for positive values of A. At zero there is a jump-discontinuity describing the probability for zero WTP.

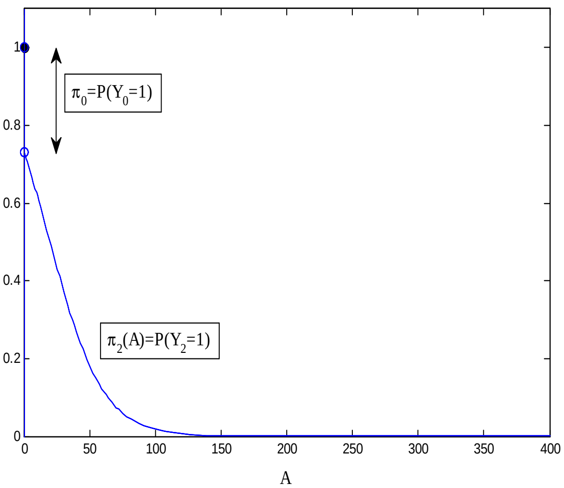


Figure 1: An illustration of the trinomial spike model with $\beta_0 = 1$ and $\beta_1 = -0.05$.

The model has two parameters, β_0, and β_1, and it is assymetric with zero probability for rejecting a negative bid. An illustration of a trinomial spike model is shown in Figure 1. The parameter β_0 is directly related to the probability of rejecting any positive cost (see 2) whereas the parameter β_1 describes the steepness of the curve. A positive (negative) β_0 implies that more (less) than 50 % has zero WTP, that is rejects any positive cost. For the parameter β_1, a larger absolute value implies a steeper slope of the response curve.

The mean of the trinomial spike model is

$$\mu = -ln(1 + e^{\beta_0})/\beta_1$$

and the median is

$$\rho = \left\{ \begin{array}{cc} 0, & \text{if } \beta_0 < 0 \\ -\beta_0/\beta_1, & \text{if } \beta_0 \geq 0 \end{array} \right\}.$$

3 Estimation of the trinomial spike model

Assume that we have independent observations on N individuals consisting of the bid A_i presented to the i:th individual, an indicator of zero WTP, y_{0i}, and an indicator of whether or not the individual accepts the bid, y_{2i}. The maximum likelihood parameter estimates are then found by maximizing the log likelihood

$$\ln L(\theta) = \sum_{i=1}^{N} y_{0i} \ln \pi_0 + (1 - y_{0i} - y_{2i}) \ln (1 - \pi_0 - \pi_2(A_i)) + y_{2i} \ln \pi_2(A_i)$$

Using the chain rule the partial derivatives with respect to the parameters are given by

$$\left(\frac{\partial \pi(A_i)}{\partial \theta}\right)^T \left(\frac{\partial l_i}{\partial \pi(A_i)}\right)$$

where l_i is the $i : th$ term of the log likelihood, $\pi(A_i) = (\pi_0,\ \pi_2(A_i))^T$ and $\theta = (\beta_0,\ \beta_1)^T$. The score function becomes

$$U(\theta) = \sum_{i=1}^{N} \left(\frac{\partial \pi(A_i)}{\partial \theta}\right)^T \left(\frac{\partial l_i}{\partial \pi(A_i)}\right).$$

The parameter estimates are obtained from the system of likelihood equations $U\left(\widehat{\theta}\right) = 0$. These equations are nonlinear in the parameters and the solutions have to be found numerically. Here we use the method of scoring which is an iterative scheme for succesively updating the parameter estimates as follows

$$\widehat{\theta}^{(s+1)} = \widehat{\theta}^{(s)} + M\left(\widehat{\theta}^{(s)}\right)^{-1} U\left(\widehat{\theta}^{(s)}\right).$$

The method of scoring is essentially a Newton-Raphson type scheme which uses the Fisher information matrix as an approximation of the Hessian matrix of second order derivatives. The Fisher information matrix for the trinomial model is

$$M(\theta) = \sum_{i=1}^{N} D_i^T V_i^{-1} D_i$$

where

$$D_i = \frac{\partial \pi(A_i)}{\partial \theta} = \begin{pmatrix} -\pi_0(1-\pi_0) & 0 \\ \pi_2(A_i)(1-\pi_2(A_i)) & A_i \pi_2(A_i)(1-\pi_2(A_i)) \end{pmatrix}$$

and

$$V_i = cov\begin{pmatrix} Y_{0i} \\ Y_{2i} \end{pmatrix} = \begin{pmatrix} \pi_0\,(1-\pi_0) & -\pi_0\pi_2(A_i) \\ -\pi_0\pi_2(A_i) & \pi_2(A_i)\,(1-\pi_2(A_i)) \end{pmatrix}.$$

3.1 Example: Environmentally friendly clothes data

Optimal designs for trinomial spike models require knowledge of the true model parameters. This knowledge can come from previous studies, pilot studies, theoretical considerations etc. We use data from a pilot study: a CVE in which the purpose was to estimate WTP for environment friendly production of clothes (see Levinson, 2010). The estimated model parameters are then used as initial guesses of the true parameters. In this experiment, respondents were asked whether they would accept a specified additional cost A_i for a top (or t-shirt) that had been ecologically produced using organic cotton, in comparison to an ordinary top (or t-shirt). There were three different bid levels used in the experiment. In addition, they were asked whether they would accept *any* additonal cost at all. Putting all respondents with zero WTP in the first point, we have the following data:

$$\begin{aligned} A &= \{\ 40 \quad 60 \quad 80\ \} \\ N &= \{\ 179 \quad 108 \quad 104\ \} \\ y_0 &= \{\ 73 \quad 0 \quad 0\ \} \\ y_2 &= \{\ 86 \quad 67 \quad 47\ \} \end{aligned}$$

We can see that 73 respondents (18.7%) had 0 WTP, that is they were not willing to accept any additional cost for the top, whereas 318 $(179-73+108+104)$ had positive WTP. Out of the 106 $(179-73)$ who were asked if they accepted 40 SEK specifically, 86 accepted that amount. Similarly, 67 out of 108 and 47 out of 104 accepted an additional cost of 60 SEK and 80 SEK respectively.

Using the methods from above and these data, we estimate the parameters of the trinomial spike model, $\theta = (\beta_0, \beta_1)^T$ to be $\widehat{\theta} = (1.4849, -0.0242)^T$ with estimated variances $\widehat{V}(\widehat{\beta}_0) = 0.12937$ and $\widehat{V}(\widehat{\beta}_1) = 0.0020345$ obtained from the diagonal of $M\left(\widehat{\theta}\right)^{-1}$. A model with these parameter values is shown in figure 2. We can see that the estimated probability of 0 WTP is around

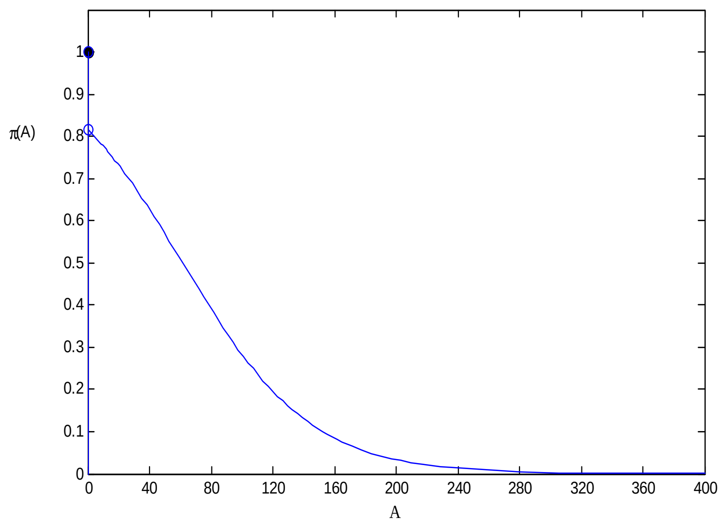


Figure 2: An illustration of the trinomial spike model with $\beta_0 = 1.4849$ and $\beta_1 = -0.0242$.

20% (which is close to the observed proportion), and that the estimated probability of $WTP \geq 280$ is practically 0. The mean WTP is 70 and the median WTP is 61.

Using these estimates from the CVE data as initial guesses of the true parameter values of the trinomial spike model, we now find optimal designs for this CVE, i.e. find the optimal number of bids, the size of these bids, and the proportion of respondents to allocate to each bid.

4 Optimal design

In planning experiments we need to specify n bids, or design points; $A_1, A_2, ..., A_n$, as well as the proportion of respondents to allocate to each bid; $\omega_1, \omega_2, ..., \omega_n$. This design can be written

$$\xi = \left\{ \begin{array}{cccc} A_1 & A_2 & ... & A_n \\ \omega_1 & \omega_2 & ... & \omega_n \end{array} \right\}.$$

The precision with which the parameter vector is estimated is a function of the number of bids, the size of the bids, and the number of respondents allocated to each bid. The optimal number of bids as well as the proportion of respondents allocated to each bid depend, in turn, on the true population parameter vector θ. In a trinomial spike model $\theta = (\beta_0, \beta_1)^T,\ \theta \in \Theta$. Because this vector is (most of the time) unknown, researchers need to provide guesses. If a single guess for the parameter values is available, a locally optimal design can be found, which is the best design given that the guessed values are true.

Locally optimal designs

The first step is to decide on a criterion that is relevant for the experiement. Two criteria are examined here. One is the D-optimality criterion which is appropriate if we want to achieve high precision in the estimation of the model parameters. The other criterion is that of c-optimality which is appropriate when the main focus of the study is estimation of some function of the model parameters, such as median WTP.

D-optimality

A D-optimal design minimizes the criterion function

$$\ln \det M(\xi, \theta)^{-1},$$

where $M(\xi, \theta)$ is the information matrix of the design ξ. The rationale for using this criterion function is that it gives minimal generalized variance, and thereby also the smallest possible asymptotic confidence region for the parameters.

In order to check whether a candidate design is optimal or not one can use a plot of the standardized predictor variance $d(A, \xi, \theta)$. For the trinomial spike model it is given by

$$d(A, \xi, \theta) = tr\left[m(A, \theta)\, M(\xi, \theta)^{-1}\right],$$

where $m(A, \theta)$ is the information from a one-point design at A. From the General Equivalence Theorem (Kiefer, 1959; Kiefer and Wolfowitz, 1959) it follows that a design ξ^* is D-optimal if and only if one of the following equivalent conditions hold

1. *The design* ξ^* *minimizes* $\det M(\xi, \theta)^{-1}$
2. *The maximum of* $d(A, \xi^*, \theta) \leq 2$
3. $d(A, \xi^*, \theta)$ *achieves its maxima at the points of the design*

Often the optimality of the design can easily be verified or disproved by a graphical examination of the plot of $d(A, \xi, \theta)$. An example of such a plot is given in Figure 3 and will be discussed below. Examples of D-optimal designs for various parameter sets are shown in Table 1. It can be seen that there is a strong dependence on the parameters when it comes to the location of the design points as well as the number of design points. For example, the one point D-optimal design is at $A = 66$ for $(\beta_0, \beta_1)^T = (2, -0.05)^T$ which can be contrasted to the one point design at $A = 234$ for $(\beta_0, \beta_1)^T = (0.4, -0.01)^T$. Note however that all respondents are asked if they are willing to accept *any* positive cost in addition to being asked if they accept the bid A. That is, there are effectively two design points in these designs. Therefore, singularity is not an issue with the one-point designs here. We can however see that when the probability of zero WTP, π_0, is close to zero the D-optimal design consists of two design points.

For the environmentally friendly clothes data the parameter estimates were found to be $\widehat{\theta} = (1.4849, \ -0.0242)^T$. Assuming these estimates are the true model parameters the following one point design is locally D-optimal

$$\xi^* = \left\{ \begin{array}{c} 122 \\ 1 \end{array} \right\}.$$

We can see that this design is D-optimal by examining Figure 3 because it has one maximum at the design point, and this maximum is not larger than 2. Given these parameters it is optimal to ask all individuals whether or not they are willing to accept an additional cost of 122 SEK for the top in addition to if they are willing to accept *any* additional cost at all.

Table 1: Locally D-optimal designs for some parameter sets.

D									
$\binom{\beta_0}{\beta_1}$	$\binom{-8}{-0.05}$	$\binom{-4}{-0.05}$	$\binom{-2}{-0.05}$	$\binom{-1}{-0.05}$	$\binom{0}{-0.05}$	$\binom{1}{-0.05}$	$\binom{2}{-0.05}$	$\binom{4}{-0.05}$	$\binom{8}{-0.05}$
π_0	≈ 1	0.98	0.88	0.73	0.50	0.27	0.12	0.02	≈ 0
ρ	0	0	0	0	0	20	40	80	160
ξ^*	$\left\{\begin{matrix} 32 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 32 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 34 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 37 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 43 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 53 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 66 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 63 & 106 \\ 0.33 & 0.67 \end{matrix}\right\}$	$\left\{\begin{matrix} 130 & 191 \\ 0.49 & 0.51 \end{matrix}\right\}$
$\binom{\beta_0}{\beta_1}$	$\binom{-1.6}{-0.01}$	$\binom{-0.8}{-0.01}$	$\binom{-0.4}{-0.01}$	$\binom{-0.2}{-0.01}$	$\binom{0}{-0.01}$	$\binom{0.2}{-0.01}$	$\binom{0.4}{-0.01}$	$\binom{0.8}{-0.01}$	$\binom{1.6}{-0.01}$
π_0	0.83	0.69	0.60	0.55	0.50	0.45	0.40	0.31	0.17
ρ	0	0	0	0	0	20	40	80	160
ξ^*	$\left\{\begin{matrix} 175 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 191 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 202 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 209 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 217 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 225 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 234 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 254 \\ 1 \end{matrix}\right\}$	$\left\{\begin{matrix} 302 \\ 1 \end{matrix}\right\}$

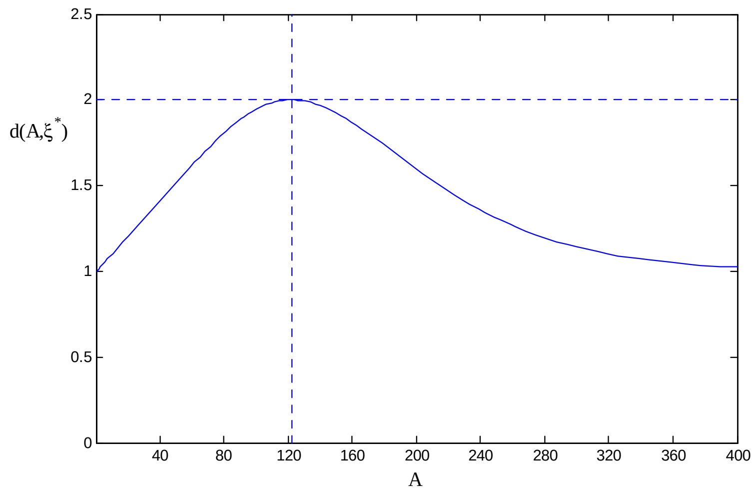


Figure 3: The standardized predictor variance for the D-optimal design when $\theta = (\beta_0, \beta_1)^T = (1.4849, -0.0242)^T$.

c-optimality

Sometimes the objective is to estimate just one parameter or some function $g(\theta)$ of the parameters as precisely as possible, for instance when the primary interest of the study is to estimate median WTP rather than the complete response curve. A c-optimal design minimizes the approximate variance of $g\left(\widehat{\theta}\right)$

$$c^T M(\xi, \theta)^{-1} c$$

where c is a $p-$vector. If the funtion $g(\theta)$ is nonlinear in θ, the vector $c = \frac{\partial g(\theta)}{\partial \theta}$. For example, when estimation of ρ, median WTP, is the focus

$$g(\theta) = -\frac{\beta_0}{\beta_1}, \quad \beta_0 \geq 0$$

and

$$c^T = \left(-\frac{1}{\beta_1}\ ,\ \frac{\beta_0}{\beta_1^2}\right)^T.$$

The standardized predictor variance is now

$$d_c(A, \xi, \theta) = tr\left[m(A, \theta) M(\xi, \theta)^{-1} c \left(c^T M(\xi, \theta)^{-1} c\right) c^T M(\xi, \theta)^{-1}\right]$$

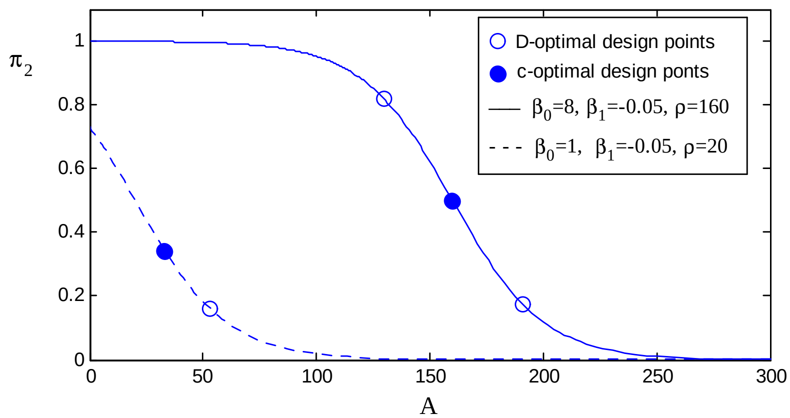


Figure 4: D- and c-optimal design points for two response curves.

and the analogous conditions for c-optimality are

1. *The design* ξ^* *minimizes* $c^T M(\xi, \theta)^{-1} c$
2. *The maximum of* $d_c(A, \xi^*, \theta) \leq 1$
3. $d_c(A, \xi^*, \theta)$ *achieves its maxima at the points of the design.*

Table 2 gives examples of c-optimal designs for estimation of median WTP for the same parameter sets as in Table 1. Practically all c-optimal designs are different from the corresponding D-optimal designs. For those cases where both are one-point designs the design point is closer to ρ in the c-optimal design. Exactly how close to ρ the design point is, is also determined by the parameter β_1. A smaller absolute value of β_1 gives a less steep slope which implies a response curve that is spread over a wider range of values and the design point is farther from ρ. Figure 4 compares the D- and c-optimal design points for two response curves. The solid line gives an example of a response curve where almost everyone has a positive WTP, that is π_0 is close to zero. The D-optimal design requires 2 design points (in addition to the zero point) while one is sufficient for the c-optimal design. The design point at zero does not provide much information in this case due to the high proportion of positive WTP. For the other response curve, both designs consist of one point, although at different positions.

Table 2: Locally c-optimal designs for estimation of median WTP μ.

c									
$\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$	$\begin{pmatrix}-8\\ -0.05\end{pmatrix}$	$\begin{pmatrix}-4\\ -0.05\end{pmatrix}$	$\begin{pmatrix}-2\\ -0.05\end{pmatrix}$	$\begin{pmatrix}-1\\ -0.05\end{pmatrix}$	$\begin{pmatrix}0\\ -0.05\end{pmatrix}$	$\begin{pmatrix}1\\ -0.05\end{pmatrix}$	$\begin{pmatrix}2\\ -0.05\end{pmatrix}$	$\begin{pmatrix}4\\ -0.05\end{pmatrix}$	$\begin{pmatrix}8\\ -0.05\end{pmatrix}$
π_0	≈ 1	0.98	0.88	0.73	0.50	0.27	0.12	0.02	≈ 0
ρ	0	0	0	0	0	20	40	80	160
ξ^*	$\left\{\begin{matrix}32\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}32\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}35\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}41\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}10 & 81\\ 0.58 & 0.42\end{matrix}\right\}$	$\left\{\begin{matrix}33\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}49\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}82\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}160\\ 1\end{matrix}\right\}$
$\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$	$\begin{pmatrix}-1.6\\ -0.01\end{pmatrix}$	$\begin{pmatrix}-0.8\\ -0.01\end{pmatrix}$	$\begin{pmatrix}-0.4\\ -0.01\end{pmatrix}$	$\begin{pmatrix}-0.2\\ -0.01\end{pmatrix}$	$\begin{pmatrix}0\\ -0.01\end{pmatrix}$	$\begin{pmatrix}0.2\\ -0.01\end{pmatrix}$	$\begin{pmatrix}0.4\\ -0.01\end{pmatrix}$	$\begin{pmatrix}0.8\\ -0.01\end{pmatrix}$	$\begin{pmatrix}1.6\\ -0.01\end{pmatrix}$
π_0	0.83	0.69	0.60	0.55	0.50	0.45	0.40	0.31	0.17
ρ	0	0	0	0	0	20	40	80	160
ξ^*	$\left\{\begin{matrix}182\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}222\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}280\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}346\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}49 & 404\\ 0.58 & 0.42\end{matrix}\right\}$	$\left\{\begin{matrix}73\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}106\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}149\\ 1\end{matrix}\right\}$	$\left\{\begin{matrix}212\\ 1\end{matrix}\right\}$

5 The minimax approach

It is often difficult to guess specific values of the true parameters. Furthermore, locally optimal designs may not be robust to erroneous guesses. In a minimax approach we guess a parameter subspace Θ_0, that belongs to the parameter space Θ. The subset Θ_0 is here defined as a rectangular region

$$\Theta_0 = \left[\beta_0^L, \beta_0^U\right] \times \left[\beta_1^L, \beta_1^U\right]$$

where the upper $\left(\beta_0^U,\ \beta_1^U\right)$ and lower $\left(\beta_0^L,\ \beta_1^L\right)$ limits of the model parameters are selected by the researcher. The minimax optimal design is then found as the design that is associated with the least favorable scenario, in terms of the chosen criterion function.

A minimax optimal design is found as the minimum over the design space of

$$\Psi(\xi) = \max_{\theta \in \Theta_0} \psi \left\{M(\xi, \theta)\right\}$$

where $\psi \left\{M(\xi, \theta)\right\}$ is the criterion function. A minimax D-optimal design is obtained by minimizing

$$\Psi(\xi) = \max_{\theta \in \Theta_0} \ln \det M\left(\xi, \theta\right)^{-1}$$

and $\Psi(\xi)$ is thus the largest possible generalized variance if θ belongs to Θ_0. Accordingly a minimax c-optimal design is obtained by minimizing

$$\Psi(\xi) = \max_{\theta \in \Theta_0} c^T M\left(\xi, \theta\right)^{-1} c.$$

The minimax design provides a protection against the worst that can happen as long as the true parameter vector θ is in Θ_0, which is a desirable robustness property. The price may be reduced efficiency for some paramter values. In reality, especially when prior information about the parameters is insufficient or unreliable, the advantages may exceed the costs. Efficiency evaluations are performed for some examples in section 5.5.

5.1 The H algorithm

It is often mathematically and numerically difficult to find a minimax design. However, a relation between minimax designs and so called optimum on average designs can be used to facilitate the search. An optimum on average design is based on a prior distribution $\pi(\theta)$ of the parameters. It minimizes the expected value of the criterion function taken over the prior distribution of θ

$$E\theta \left[\psi \left\{M(\xi, \theta)\right\}\right].$$

According to the General Equivalence Theorem the equivalent conditions for an optimum on average design ξ^* are, similarly as before, that the design ξ^* minimizes $E\theta \left[\psi \left\{M(\xi, \theta)\right\}\right]$, that the expectation of the standardized predictor variance

$$d_B(A, \xi^*) = E\theta \left[d(A, \xi^*, \theta\right] \leq s, \tag{3}$$

and that $d_B(A, \xi^*)$ achieves its maxima at the design points. We have $s = 2$ for D-optimality and $s = 1$ for c-optimality.

Nyquist and Häggström (1998) show that if $\pi^*(\theta)$ is a prior distribution for θ with support only at points $\theta \in \Theta_0$, and if ξ^* is an optimum on average design with respect to $\pi^*(\theta)$ such that $\psi \left\{M(\xi, \theta)\right\} \leq E\theta \left[\psi \left\{M(\xi, \theta)\right\}\right]$ for all $\theta \in \Theta_0$, then under general regularity conditions, ξ^* is a minimax design with respect to Θ_0. This changes the problem of finding a minimax design to the problem of finding the prior distribution $\pi^*(\theta)$, which is the *least favorable distribution* (l.f.d.), and the associated optimum on average design. Nyquist and Häggström (1998) show that the prior distribution $\pi^*(\theta)$ must satisfy

$$\pi^*(\theta) \left\{ \begin{array}{l} = 0, \text{ if } \psi \left\{M(\xi^*, \theta\right\} < \psi^* \\ \geq 0, \text{ if } \psi \left\{M(\xi^*, \theta\right\} = \psi^* \end{array} \right\}, \tag{4}$$

where $\psi^* = \max\limits_{\theta \in \Theta_0} \psi \left\{M(\xi, \theta\right\}.$

In order to find a minimax design, we can thus, for an initial prior distribution $\pi_1(\theta)$, compute the optimum on average design ξ_1, for example using the V-algorithm of Fedorov (1972). Then, $\psi \left\{M(\xi_1, \theta\right\}$ is computed and a new prior distribution, $\pi_2(\theta)$ is constructed by adding points to the previous prior $\pi_1(\theta)$ where $\psi \left\{M(\xi_1, \theta\right\}$ has its maxima. By using $\pi_2(\theta)$ a new design ξ_2 is then found and iterations continue until the requirements in (3) and (4) are satisfied. This algorithm will be illustrated using the environmentally friendly clothes data presented in section 3.1.

5.2 Minimax D-optimal design

As noted previously, the locally D-optimal design

$$\xi = \left\{ \begin{array}{c} 122 \\ 1 \end{array} \right\}$$

is optimal given that $\theta = (1.4849,\ -0.0242)^T$. However, if we are uncertain that these parameter values are close to the correct values, a better approach is to guess intervals around one or both of them. If we assume that θ belongs to $\Theta_0 = \{1.5\} \times [-0.0375, -0.0187]$, for example, we can use the H-algorithm to find the D-optimal minimax design. We choose an initial prior distribution, for example the distribution that assigns 0.5 point mass at $\beta_1 = -0.0375$ and 0.5 point mass at $\beta_1 = -0.0187$. We denote this distribution by

$$\pi_1(\beta_1) = \left\{ \begin{array}{cc} -0.0375 & -0.0187 \\ 0.5 & 0.5 \end{array} \right\}$$

Using the V-algorithm to find the optimum on average design for this prior distribution yields the initial design

$$\xi_1 = \left\{ \begin{array}{c} 103 \\ 1 \end{array} \right\}.$$

Figure 5 shows a graph of $\psi\{M(\xi_1, \theta)\}$ in which $E\theta\left[\psi\{M(\xi_1, \theta)\}\right] = -5.469$ (indicated by the horizontal line in the figure). To see that $\pi_1(\beta_1)$ is not a *l.f.d.*, and that ξ_1 is not a minimax design, we note that the condition $\psi\{M(\xi_1, \theta)\} \leq E\theta\left[\psi\{M(\xi_1, \theta)\}\right]$ for all $\theta \in \Theta_0$ does not hold. The maxima indicates that mass should be added to the prior distribution at $\beta_1 = -0.0375$.

A new prior distribution

$$\pi_2(\beta_1) = \left\{ \begin{array}{cc} -0.0375 & -0.0187 \\ 0.51 & 0.49 \end{array} \right\}$$

yields the new optimum on average design

$$\xi_2 = \left\{ \begin{array}{c} 102 \\ 1 \end{array} \right\}$$

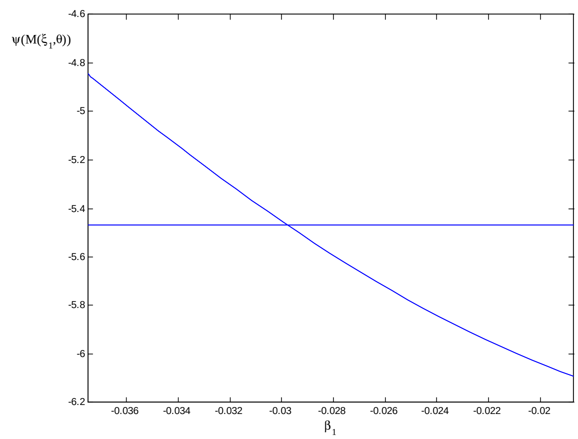


Figure 5: The criterion function evaluated for the initial design ξ_1

which, after checking, is not minimax. Continuing to add mass to the point $\beta_1 = -0.0375$, after several iterations, finally gives us the prior distribution

$$\pi^*(\beta_1) = \left\{ \begin{array}{c} -0.0375 \\ 1 \end{array} \right\}$$

with the corresponding optimum on average design

$$\xi^* = \left\{ \begin{array}{c} 79 \\ 1 \end{array} \right\}.$$

Figure 6 shows $\psi\left\{M(\xi^*, \theta)\right\}$ for all $\theta \in \Theta_0$. Now $E\theta\left[\psi\left\{M(\xi_1, \theta)\right\}\right] = -5.0099$ is in the point $\beta_1 = -0.0375$ which is also the maximum in the interval. This indicates that $\pi^*(\theta)$ is a *l.f.d.* and that the design ξ^* is a minimax D-optimal design.

D-optimal minimax designs for varying regions Θ_0 are shown in Table 3. For example, the minimax design (and also locally D optimal design) for $\theta = (1.5, -0.02)^T$ is $\xi = \left\{ \begin{array}{c} 147 \\ 1 \end{array} \right\}$ whereas the minimax design for θ belonging to $\Theta_0 = [1, 2] \times [-0.1, -0.001]$ is $\xi = \left\{ \begin{array}{c} 30 \\ 1 \end{array} \right\}$. We can see that all minimax designs for the specified intervals are one-point designs (in practice two-

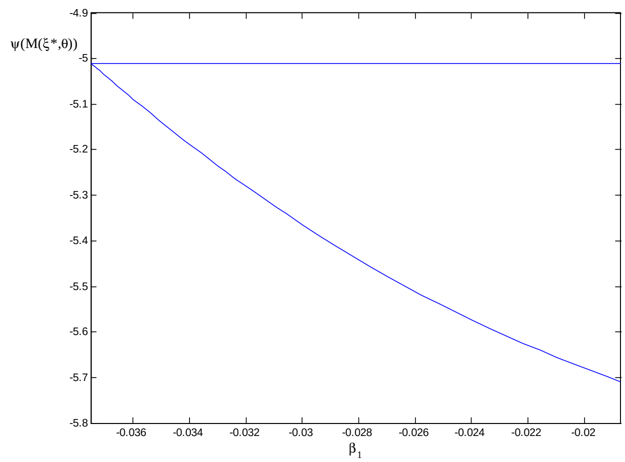


Figure 6: The criterion function evaluated for the minimax design ξ^*

point designs as mentioned earlier) and that they are substantively different depending on the parameter regions.

As noted previously, the *l.f.d.* for $\theta \in \Theta_0 = \{1.5\} \times [-0.0375, -0.0187]$, is unit mass for $\theta = (1.5, -0.0375)^T$. This is the logistic curve with the steepest slope within the region. The same holds for all regions in table 3. Table 7 shows pictures of *l.f.d.'s* for the regions in table 3. For example, the *l.f.d.* for $\theta \in \Theta_0 = [-1, 2] \times [-0.1, -0.001]$, illustrated by the fourth figure in the last row, is unit mass for $\theta = (-1, -0.1)^T$ whereas the *l.f.d.* for $\theta \in \Theta_0 = [1, 2] \times [-0.1, -0.001]$, illustrated by the fifth figure in the last row, is 0.5 mass for $\theta = (1, -0.1)^T$ and 0.5 mass for $\theta = (2, -0.1)^T$. As shown, unit mass is assigned to the curves with the steepest slopes. The pattern for β_0 is not as straightforward.

Table 5 shows the *l.f.d.'s* for varying intervals of β_0 such that $\Theta_0 = \left[\beta_0^L \leq \beta_0 \leq \beta_0^U\right] \times [-0.1 \leq \beta_1 \leq -0.001]$. We can see that mass is always in one or more of the four corners of the subspace Θ_0.When $\beta_0^L = -1$, unit mass is in the lower endpoint of β_0 (-1), regardless of the upper endpoint. This is also the case for $\beta_0^L = 0$. However when $\beta_0^L = 0.5$, the mass in the lower endpoint of β_0 decreases as the upper endpoint increases, and for $\beta_0^U = 2$, the mass is 0.9 and 0.1 in the two endpoints respectively. The same pattern holds for $\beta_0^L = 1$ with increasing mass at the upper endpoint as the interval

Table 3: Minimax D-optimal designs for some parameter intervals.

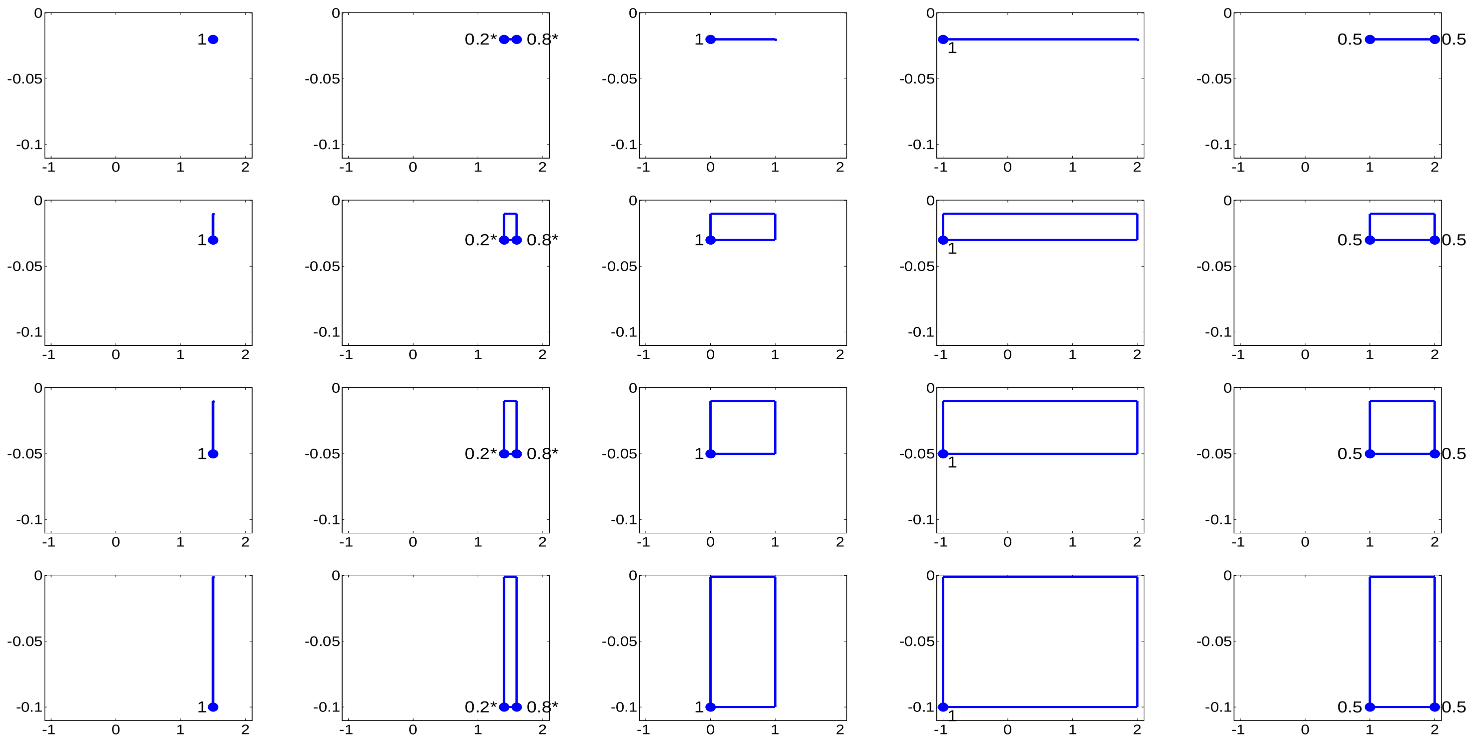
$[\beta_1^L, \beta_1^U]$	$[\beta_0^L, \beta_0^U]$ $[1.5, 1.5]$	$[1.4, 1.6]$	$[0, 1]$	$[-1, 2]$	$[1, 2]$
$[-0.02, -0.02]$	$\left\{\begin{array}{c}147\\1\end{array}\right\}$	$\left\{\begin{array}{c}150\\1\end{array}\right\}$	$\left\{\begin{array}{c}108\\1\end{array}\right\}$	$\left\{\begin{array}{c}93\\1\end{array}\right\}$	$\left\{\begin{array}{c}150\\1\end{array}\right\}$
$[-0.03, -0.01]$	$\left\{\begin{array}{c}97\\1\end{array}\right\}$	$\left\{\begin{array}{c}101\\1\end{array}\right\}$	$\left\{\begin{array}{c}73\\1\end{array}\right\}$	$\left\{\begin{array}{c}62\\1\end{array}\right\}$	$\left\{\begin{array}{c}100\\1\end{array}\right\}$
$[-0.05, -0.01]$	$\left\{\begin{array}{c}60\\1\end{array}\right\}$	$\left\{\begin{array}{c}60\\1\end{array}\right\}$	$\left\{\begin{array}{c}43\\1\end{array}\right\}$	$\left\{\begin{array}{c}37\\1\end{array}\right\}$	$\left\{\begin{array}{c}60\\1\end{array}\right\}$
$[-0.1, -0.001]$	$\left\{\begin{array}{c}30\\1\end{array}\right\}$	$\left\{\begin{array}{c}30\\1\end{array}\right\}$	$\left\{\begin{array}{c}22\\1\end{array}\right\}$	$\left\{\begin{array}{c}19\\1\end{array}\right\}$	$\left\{\begin{array}{c}30\\1\end{array}\right\}$

length increases. For $1.4 \leq \beta_0 \leq 1.6$, the mass is 0.2 and 0.8 respectively in the two endpoints, and the mass in the upper endpoint increases as the interval length increases such that $1.4 \leq \beta_0 \leq 2$. Least favorable distributions for intervals with $\beta_0^U > 2$ are more complicated to find. In the search for these distributions, mass was added to other points than the corners of the rectangular parameter space. Patterns outside the studied intervals are left for future research.

5.3 Minimax c-optimal design

The minimax c-optimal designs are also derived following the H algorihm as described in the previous section. The resulting designs are reported in Table 6. Due to numerical issues when β_1 is close to zero the largest interval for β_1 is cut off at -0.01 here. These are either one or two point designs depending on the size of the parameter regions. The corresponding *l.f.d.'s* are found in Table 7 . It can be noted that these are all different from the D-optimal designs (as for the locally optimal designs). Furthermore, the *l.f.d's* often have masses in other corners of the parameter space. In a couple of cases the

Table 4: Least favourable distributions for the minimax designs in Table 3. Lines indicate the parameter space and points with mass are indicated with circles.



*The l.f.d. is also unit mass for the point in which $\beta_0 = 1.6$.

Table 5: Least favorable distributions for $\theta \in \Theta_0 = \left[\beta_0^L \leq \beta_0 \leq \beta_0^U\right] \times [-0.1 \leq \beta_1 \leq -0.001]$

	β_0^L				
β_0^U	-1	0	0.5	1	1.4
2	$\left\{\begin{matrix}(-1,-0.1)\\1\end{matrix}\right\}$	$\left\{\begin{matrix}(0,-0.1)\\1\end{matrix}\right\}$	$\left\{\begin{matrix}(0.5,-0.1) & (2,-0.1)\\0.9 & 0.1\end{matrix}\right\}$	$\left\{\begin{matrix}(1,-0.1) & (2,-0.1)\\0.5 & 0.5\end{matrix}\right\}$	$\left\{\begin{matrix}(2,-0.1)\\1\end{matrix}\right\}$
1.6	$\left\{\begin{matrix}(-1,-0.1)\\1\end{matrix}\right\}$	$\left\{\begin{matrix}(0,-0.1)\\1\end{matrix}\right\}$	$\left\{\begin{matrix}(0.5,-0.1)\\1\end{matrix}\right\}$	$\left\{\begin{matrix}(1,-0.1)\\1\end{matrix}\right\}$	$\left\{\begin{matrix}(1.4,-0.1) & (1.6,-0.1)\\0..2^* & 0.8^*\end{matrix}\right\}$
1	$\left\{\begin{matrix}(-1,-0.1)\\1\end{matrix}\right\}$	$\left\{\begin{matrix}(0,-0.1)\\1\end{matrix}\right\}$	$\left\{\begin{matrix}(0.5,-0.1)\\1\end{matrix}\right\}$		
0	$\left\{\begin{matrix}(-1,-0.1)\\1\end{matrix}\right\}$				

*The *l.f.d.* is also unit mass for the point in which $\beta_0 = 1.6$.

Table 6: Minimax c-optimal designs for some parameter intervals.

$\left[\beta_1^L,\beta_1^U\right]$	$\left[\beta_0^L,\beta_0^U\right]$ [1.5, 1.5]	[1.4, 1.6]	[0, 1]	[−1, 2]	[1, 2]
[−0.02, −0.02]	$\left\{\begin{matrix}102\\1\end{matrix}\right\}$	$\left\{\begin{matrix}107\\1\end{matrix}\right\}$	$\left\{\begin{matrix}42 & 140\\0.52 & 0.48\end{matrix}\right\}$	$\left\{\begin{matrix}110\\1\end{matrix}\right\}$	$\left\{\begin{matrix}120\\1\end{matrix}\right\}$
[−0.03, −0.01]	$\left\{\begin{matrix}205\\1\end{matrix}\right\}$	$\left\{\begin{matrix}209\\1\end{matrix}\right\}$	$\left\{\begin{matrix}85 & 290\\0.52 & 0.48\end{matrix}\right\}$	$\left\{\begin{matrix}210\\1\end{matrix}\right\}$	$\left\{\begin{matrix}243\\1\end{matrix}\right\}$
[−0.05, −0.01]	$\left\{\begin{matrix}70 & 210\\0.022 & 0.978\end{matrix}\right\}$	$\left\{\begin{matrix}70 & 210\\0.0017 & 0.983\end{matrix}\right\}$	$\left\{\begin{matrix}85 & 290\\0.52 & 0.48\end{matrix}\right\}$	$\left\{\begin{matrix}50 & 210\\0.015 & 0.985\end{matrix}\right\}$	$\left\{\begin{matrix}243\\1\end{matrix}\right\}$
[−0.1, −0.01]	$\left\{\begin{matrix}30 & 210\\0.02 & 0.98\end{matrix}\right\}$	$\left\{\begin{matrix}30 & 210\\0.0026 & 0.974\end{matrix}\right\}$	$\left\{\begin{matrix}85 & 290\\0.52 & 0.48\end{matrix}\right\}$	$\left\{\begin{matrix}50 & 210\\0.014 & 0.986\end{matrix}\right\}$	$\left\{\begin{matrix}243\\1\end{matrix}\right\}$

l.f.d. also has support on a point away from the corner on the edge of the parameter space.

5.4 Efficiency of the Minimax Designs

As noted, substantively different design points will be optimal given different true parameter values. In addition, locally optimal designs may not be robust to erronous guesses far away from the true parameters. The D-efficiency of a design ξ compared to the locally optimal design ξ^* is

$$\left(\frac{\det M(\xi^*,\theta)^{-1}}{\det M(\xi,\theta)^{-1}}\right)^{\frac{1}{2}}$$

for the trinomial model. D-efficencies close to 1 imply that the design ξ is nearly as good as ξ^*. If the efficency is lower it means that more observations are needed to achive the same precision as ξ^*. For example, an efficency of 0.5 would imply that twice as many observations are required for ξ to

Table 7: Least favourable distributions for the minimax designs in Table

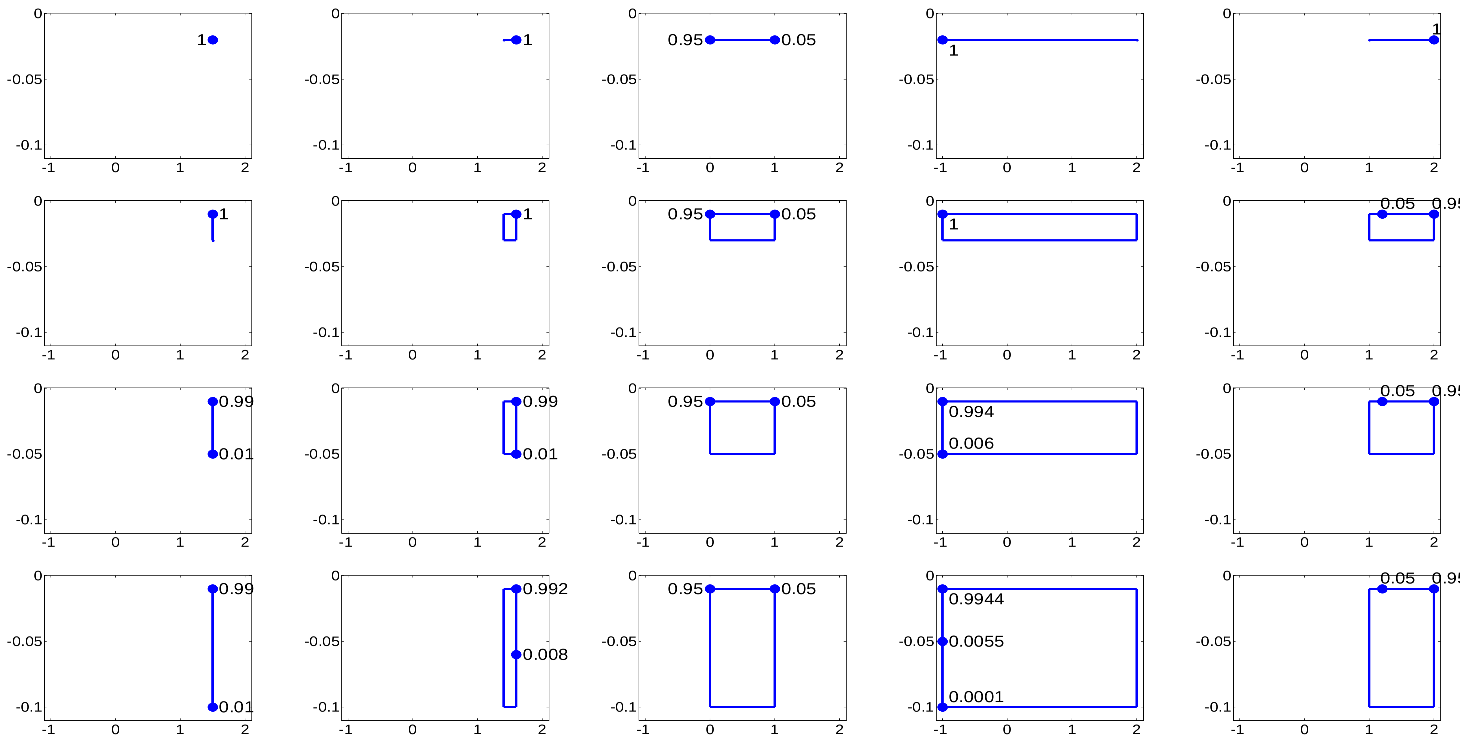


Table 8: D-efficiencies of the minimax D-optimal design based on the largest uncertainty region compared to locally optimal designs for different parameter values. Corresponding D-efficiencies of the locally optimal design based on $\beta_0 = 1.5$ and $\beta_1 = -0.02$ as the true values are given in parenthesis..

	β_0				
β_1	-1	0	1	1.5	2
-0.001	0.13 (**0.37**)	0.10 (**0.28**)	0.06 (**0.18**)	0.05 (**0.14**)	0.03 (**0.10**)
-0.01	0.42 (**0.98**)	0.33 (**0.94**)	0.21 (**0.81**)	0.16 (**0.71**)	0.12 (**0.50**)
-0.02	0.59 (**0.90**)	0.47 (**0.94**)	0.32 (**0.99**)	0.24 (**1**)	0.18 (**0.99**)
-0.03	**0.71** (0.64)	0.59 (**0.69**)	0.42 (**0.77**)	0.33 (**0.82**)	0.25 (**0.88**)
-0.05	**0.87** (0.24)	**0.78** (0.26)	**0.60** (0.30)	**0.49** (0.33)	**0.39** (0.36)
-0.1	**1** (0.01)	**0.99** (0.01)	**0.92** (0.01)	**0.85** (0.02)	**0.75** (0.02)

be equivalent to ξ^*. Table 8 presents D-efficiencies of the minimax design $\xi = \left\{ \begin{array}{c} 19 \\ 1 \end{array} \right\}$ for $\Theta_0 = [1, 2] \times [-0.1, -0.001]$, which represents the maximal uncertainty region of those in Table 3, when the true parameters are varied within this region. The table also presents the D-efficiencies of the locally optimal design $\xi = \left\{ \begin{array}{c} 147 \\ 1 \end{array} \right\}$ for the estimated parameter values $\widehat{\beta}_0 \approx 1.5$ and $\widehat{\beta}_1 \approx -0.02$, representing no uncertainty about this point. As expected, the locally optimal design performs better in a region close to these estimates. The minimax design protects against the least favorable scenario within the interval, the *l.f.d.* has unit mass at the point $(-1, -0.1)$. The minimax design performs best when the response curve is steep, that is when β_1 is close to -0.1. But since the minimax design is also a one-point design it cannot be better than the locally optimal design for those parameter values. In summary when the estimate/guess is bad the minimax designs perform better, otherwise the locally optimal design is better.

6 Conclusions

When planning an experiment, a reseaarcher needs to specify the number of bids, the size of the bids, and the proportion of respondents allocated to each bid. When the situation is such that all respondents are asked whether or not they accept any positive cost in addition to whether or not they accept a certain amount, the responses can be modelled by a trinomial spike model. Optimal designs for this model are dependent on the true parameter values. They can therefore not be used in practice. However, preliminary information about the parameters is often available, e.g. from previous studies, from pilot studies, from theoretical or logical considerations etc. In addition, this preliminary information can often be presented in forms of plausable intervalls or regions for the parameters. We have demonstrated how D-optimal and c-optimal minimax designs are derived for the trinomial spike model with the help of preliminary information. We have also illustrated the methods and results using data from a CVE to estimate WTP for environment friendly production of clothes.

We have seen strong dependences on the parameters when it comes to the location of the design points as well as the number of design points for locally D- and c- optimal designs. A researcher's different guesses of the true parameters can thus result in substantively different designs, which in turn will affect the precision in estimation of the parameters. If the true parameters lie far away from the guesses, and if the logistic curve is steeper than expected, the loss in precision may be substantial. We have proposed using the minimax approach to find the optimal number of bids, the bid sizes, and proportions of respondents to allocate to each bid. The researcher guesses a region of plausible parameter values, and the approach provides protection against the worst case scenario.

Because it is often mathematically and numerically difficult to find a minimax design, we have proposed using a relation between minimax designs and optimum on average designs to facilitate the search. Instead of searching for a minimax design, we can search for the prior distribution which is the least favorable distribution and the associated optimum on average design. Our findings for D-optimal minimax designs suggest that least favorable distributions are those that places mass in the corners of the regions, and also those that places mass in points that indicate the steepest logistic curves. This should be studied more. We also noted that many of our minimax designs were one-point designs, which in practice are two-point designs because the zero-point is always present in these types of designs. However, designs for models in which the probability of zero WTP was practically zero was more difficult to find, and it seems that more design points might be needed, which should be studied further. In practice, however, if a researcher suspects that all idividuals will have positive WTP, the trinomial spike model may not be the best model to use.

We provided both D-optimal and c-optimal designs and we could see that they were not equivalent. It is therefore important for researchers to be clear about the primary interest of the study. If the primary interest is to estimate the model parameters, D-optimal designs should be used, whereas if the primary interest is to estimate the median or mean WTP for example, c-optimal designs should be used. In those cases estimation of the median is of primary interest it may also be easier for the researcher to provide guesses of the median, in addition to β_1, in which case another parameterization may be more practical. An altenative parameterization that is common for the standard logistic model is $\beta = \beta_1$ and $\mu = -\beta_0/\beta_1$. On the other hand, with the current parameterization there is a direct relationship between π_0 and the parameter β_0. Locally optimal designs remain unchanged after converting the parameter values while the minimax designs may be different because the parameter space will be changed. We note, also, that c-optimal designs for estimation of μ and ρ are not the same due to assymetry, however this is not studied here.

It is interesting to compare our results with the minimax D-optimal designs for the standard logistic model, see King and Wong (2000). Some of their minimax designs have more design points, up to 5, compared to a locally D-optimal 2-point design. One reason for that could be that their intervals are wider. Another reason could be that the other parameterization is used.

In conclusion, if we can find information in the form of plausable intervals for the parameters, using the H-algorithm as presented here, will provide us with designs that protect against the worst case (within the interval). In cases where we cannot find any reliable prior information, the methods presented here could also be useful for sequential approaches.

References

[1] Atkinson, A.C., Donev, A.N. and Tobias, R.D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, New York.

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

[3] Hanemann, W.M. (1984) Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses. *American Journal of Agricultural Economics* 66, 332-341.

[4] Hanemann, W.M. and B. Kanninen (1996) Statistical Considerations in CVM. *Valuing Environmental Preferences: Thery and Practise of the Contingent Valuation Method in the US, EU and Developing Countries.* I.J. Bateman and K.G. Willis, eds., chap. 11. Oxford: Oxford University Press.

[5] Johansson, P.O., Kriström, B. and Nyquist, H. (1993). Valuing a public good using the CVM: Bid vectors, spikes, and risk. Stockholm School of Economics, Sweden.

[6] Kiefer, J. (1959) Optimum experimental designs. *Journal of the Royal Statistical Society, Series B*, **21**, 272–319.

[7] Kiefer, J. and Wolfowitz, J. (1959) Optimum designs in regression problems. *The Annals of Mathematical Statistics*, **30**, 271–94.

[8] King, J and Wong, W-K , (2000). Minimax D-Optimal Designs for the Logistic Model, *Biometrics* 56, 1263-1267.

[9] Kriström, B. , (1997). Spike Models in Contingent Valuation, *American Journal of Agricultural Economics* **79**, 1013-1023

[10] Kriström, B., Nyquist, H. and Hanemann, W.M. (1992) Dichotomous Choice Valuation Experiments; A Mixed- distribution Approach, Mimeo, Stockholm School of Economics

[11] Levinson, E. (2010). "Green clothes" - A survey of peple's willingness to pay for environmental friendly clothes. *Master thesis*, Sustainable Enterprising, Stockholm University.

[12] Nyquist, H. and Häggström, J. (1998). Some relations between optimum Bayes designs and minimax designs. *Statistical Research Report*, 1998-6, Department of Statistics, University of Umeå.

[13] Silvey, S.D. (1980). *Optimal Design*. Chapman and Hall, London.