90-34R

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

CHOICE OF THRESHOLDS FOR EFFICIENT BINARY DISCRETE CHOICE ESTIMATION

BY

ANNA ALBERINI

AND

RICHARD T. CARSON

DISCUSSION PAPER 90-34R
OCTOBER 1993

CHOICE OF THRESHOLDS FOR EFFICIENT BINARY DISCRETE CHOICE ESTIMATION

Anna Alberini
Resources for the Future

Richard T. Carson
University of California, San Diego

September 1993

Send Comments to:

Richard T. Carson
Department of Economics, 0508
9500 Gilman Drive
University of California, San Diego
La Jolla, CA 92093-0508

Abstract

We study the loss of information in estimating the central tendency of a normal population when a discrete indicator rather than the underlying continuous response variable is used. The threshold levels which define the value taken by the discrete indicator are shown to affect the efficiency of the parameter estimates. Several methods for choosing thresholds are compared and the potential efficiency gains from optimally choosing the thresholds are shown to be large. The robustness of different design criteria to incorrect information about the parameters is explored.

1. Introduction

Consider the linear regression model:

(1) $$y_i^* = x_i\gamma + \varepsilon_i,$$

where x_i is a vector of fixed regressors and the disturbance terms ε_i are i.i.d. with expectation zero and scale parameter σ. Suppose that instead of observing the continuous response y_i^*, one observes a binary indicator variable $I\left[y_i^*;c_i\right]$ equal to 1 if y_i^* is greater than or equal to a threshold c_i and 0 otherwise.

In economics, the random variable y^* often relates to some aspect of an agent's utility level while the indicator variable typically denotes whether the agent engages in some activity of interest. In labor economics, for instance, under certain assumptions, the optimal behavior on the part of a job applicant engaging in a job search is to accept a job offer when the offered wage rate is greater than the agent's reservation wage rate. The offered wage rate takes the role of the threshold; the indicator variable takes on a value of 1 if the job applicant accepts the job and a value of 0 otherwise. Similarly, a person decides to participate in a welfare program if the financial aid provided by the program is higher than the effective income from an alternative source such as a regular job. In the market for a particular good, a buyer purchases a unit of the good when the buyer's underlying willingness to pay is higher than the threshold represented by the selling price. In a referendum on a program of public interest, the voter compares the perceived benefits with the additional tax the voter must pay if the plan is approved and votes in favor of the proposition if the former is greater. Economic theory postulates that the economic agent, when confronted with a choice among alternatives, chooses the one associated with the higher utility level. The indicator variable denotes which alternative was preferred.

We will be particularly interested in situations where the choice of the threshold(s) c_i is under the researcher's control. In marketing research, for instance, price is often manipulated, and one observes consumer response to different prices in terms of buying or not buying the good (Cameron and James, 1987b). Another area where researchers are in control of the c_i's is in contingent valuation surveys (Mitchell and Carson, 1989), an increasingly popular technique used to measure the benefits from providing environmental amenities. In this context, instead of directly eliciting a respondent's maximum willingness to pay for the proposed project, researchers are turning to a discrete choice approach that mimics a referendum voting situation. Each respondent is asked whether he would vote favorably on a proposition which, if approved, would provide the project and pass the costs on to the respondent, usually in the form of higher taxes or prices. Another significant area in which researchers have some control over the choice of the c_i's is in the large scale social experiments undertaken by the government to assess the effects of a certain policy on a sample of individuals (Ferber and Hirsch, 1982). Researchers also have direct control over the choice of the c_i's used in laboratory economic research (Smith, 1980). A less obvious instance is the elicitation of responses to common survey questions: often particular qualitative classifications, e.g., "more or less than 5", are used in lieu of continuous quantitative variables such as "duration of the unemployment spell in weeks" or "number of monthly visits to an automatic teller machine".

Even when possible, attempting to obtain the underlying y^* may not always be desirable. Several factors, including a respondent's desire to protect privacy, a diminished opportunity for strategic behavior, a reduction in the likelihood that the observations y_i^* will be contaminated by various experimental or survey response effects, and the added cost of obtaining continuous response answers, may make it more desirable to have the econometrician look at discrete data.

As in other experimental sciences, the threshold c_i can often be thought of as the "dose" of a certain treatment or stimulus that is applied to subject i. Each of the $i = 1, 2, \ldots, n$ agents in the sample can potentially be randomly assigned to a different threshold level c_i. Whether by explicitly setting a value for the threshold or by using some qualitative classification, the researcher will try to control the fractions of the sample that fall into the "one" group and the "zero" group.

Our work may be seen as an extension, in the direction of discrete choice estimation, of earlier work by econometricians (Conlisk, 1973; Aigner and Morris, 1979) on optimal design for economic experiments with continuous response variables. Our results are complementary to earlier work on choice-based sampling (Manski and Lerman, 1977; McFadden, 1981) in that, instead of designing sampling schemes and estimation strategies which exploit the known choice patterns of agents, we are influencing the choices of agents through the designation of c_i by the survey designer or experimenter. Throughout our paper, we draw heavily from the biometrics literature (Finney, 1978; Silvey, 1980) on the optimal design of dose-response experiments.

The paper is organized as follows. Section 2 illustrates the problem of estimating an unobservable random variable by means of a binary data model, introduces some notation, and provides one simple example that illustrates the effect of the thresholds on the efficiency of the estimate. Section 3 establishes an optimality result which shows that typically two threshold levels are all that are needed to minimize the variance of the estimate of a particular quantile, such as the median. We find the optimal thresholds for the median of a normal variable y^* using numerical optimization methods. Section 3 also surveys design methods and looks at the implications of different concepts of optimality on the design chosen, considers problems arising in the presence of covariates, and introduces the problem of uncertainty surrounding the true parameter values when determining the set of c_i's to use. The findings from a large simulation study that compares the different approaches to optimal experimental designs are described in Section 4. Section 5 provides some concluding remarks.

2. Discrete choice model estimation of quantiles

2.1 Notation and formulation of the problem

Assume y^* is a random variable from the location-scale family of distributions, and let μ and σ be the particular distribution's location and scale parameters: $y^* = \mu + \varepsilon$. Let F be the known cdf of ε / σ. While y^* cannot be observed directly, we assume that observations on the indicator variable y_i take on a value of 1 if the latent y_i^* is greater than the threshold c_i and 0 otherwise. The mapping from the original latent variable to the available observations can be written as $E(y_i) = \Pr(y_i = 1) = 1 - F(\alpha + \beta c_i)$, where $\alpha \equiv -\mu / \sigma$ and $\beta \equiv 1 / \sigma$.

Suppose that the econometrician is interested in estimating the 100*p*th quantile of the distribution of y^*, that is, $g = F^{-1}(p)\sigma + \mu = \dfrac{F^{-1}(p) - \alpha}{\beta}$. This quantity g may also be interpreted as the level of the effective threshold that generates an observed value of 1 in 100(1-*p*) percent of the population. Assuming that the unknown parameters $\theta = (\alpha, \beta)$ are estimated by maximum likelihood (ML) techniques and that the vector of derivatives with respect to the parameters $\partial g / \partial \theta$ has constant rank in any open neighborhood of the true parameters, the approximate variance for $\hat{g}$ can be expressed as

(2) $$Var(\hat{g}) = \frac{\partial g}{\partial \theta'} I(\theta)^{-1} \frac{\partial g}{\partial \theta},$$

where $I(\theta)$ is the Fisher information matrix for the parameters from the log likelihood for the binary discrete choice model. $I(\theta)$ is here equal to:

(3) $$I(\theta) = \sum_{i=1}^{n} \frac{f^2(\alpha + \beta c_i)}{F(\alpha + \beta c_i)[1 - F(\alpha + \beta c_i)]} \begin{bmatrix} 1 & c_i \\ c_i & c_i^2 \end{bmatrix},$$

where c_i is the threshold assigned to the *ith* sample unit.

The thresholds affect the variance of the ML estimates $\hat{\alpha}$ and $\hat{\beta}$ and enter in the determination of the variance of $\hat{g}$ (2) via the information matrix (3). Expression (3) shows that the variance also depends on both parameters of the distribution of the latent y^*. While these parameters (μ and σ) are not under the researcher's control, the thresholds are assumed to be.

An experimental design can thus be defined as the list of the thresholds c_i assigned to each sample unit (*i*=1, 2, ..., *n*). As an alternative, letting c_j, for *j* from 1 to *k*, denote the *distinct* values of the thresholds, an experimental design can also be defined as the set of pairs $\{c_j(n), n_j(n)\}$, $j = 1, 2, \ldots, k$, where $n_j(n)$ is the number of observations associated with threshold $c_j(n)$ and $\sum_{j=1}^{k} n_j(n) = n$. Assuming that the sample size n is given, the goal is to determine the thresholds and number of observations assigned to each threshold that ensure maximum efficiency of the estimate $\hat{g}$.

2.2 One simple example

The role of the thresholds can be readily grasped by considering one simple example. Consider first the model $y_i^* = \mu + \varepsilon_i$, $i = 1, 2, \ldots, n$, where each ε_i is i.i.d. normal with mean zero and known variance 1. Assume each unit in the sample is assigned to a single threshold c. The corresponding discrete choice model is the probit equation $E(y_i) = \Phi(\mu - c)$. The threshold c is absorbed into the constant term of the probit equation. The variance for the intercept term is then given by

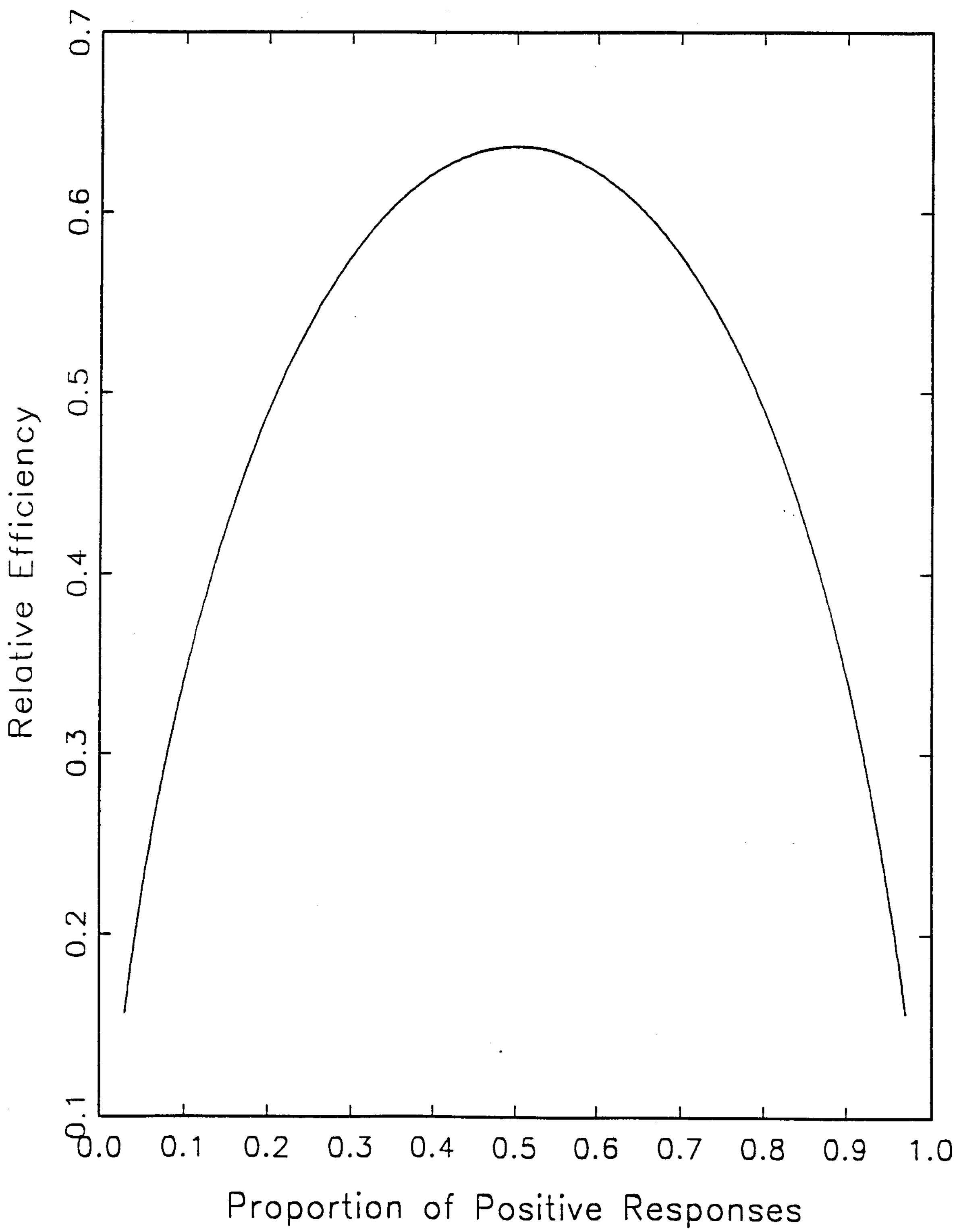
$$\text{(4)} \qquad \frac{\Phi(\mu - c)[1 - \Phi(\mu - c)]}{n\phi^2(\mu - c)}.$$

The variance of the same intercept term estimated by ML from the uncensored sample is given by $1/n$, which, divided by (4), yields the relative efficiency of the estimates for μ. The solid line in Figure 1 plots the relative efficiency for μ as a function of the split between positive responses (ones) and negative responses (zeros), the split being determined by the value of the threshold. The minimum level of the variance is achieved when $c = \mu$, which is the level of the threshold that corresponds to a 50%-50% split. Extreme splits (i.e., 90%-10%) result in much larger variance estimates for $\hat{\mu}$.

3. Optimal design criteria

Figure 1.

Relative Efficiency of the Estimate of the Median.
Probit Model with Constant Threshold c.



With the distribution of y^* belonging to a location-scale family of distributions, symmetric, and involving no covariates, we will assume that the statistic of interest g is a function of both parameters, μ and σ, of the distribution of the underlying y^*.[1] These parameters are identifiable in the binary discrete choice model if *at least two* distinct thresholds c_j are randomly assigned to the sample. Under these conditions, the threshold c_i may be treated as an additional regressor in the binary data model $E(y_i) = 1 - F(\alpha + \beta c_i)$. The regression parameter β attached to the regressor c_i is the reciprocal of the scale parameter σ of the underlying distribution, while the constant term $\hat{\alpha}$ divided by $-\hat{\beta}$ gives the estimate for μ.[2] Cameron and James (1987) apply this result to dichotomous choice contingent valuation survey data to obtain the median/mean of a normal or logistic willingness to pay (WTP).[3]

Although many of our results can be generalized in a straightforward manner to other quantiles and distributions, to keep the notation manageable throughout the remainder of this paper we will assume that our quantile of interest is the median of a normal distribution. The associated probit model is $\Pr(y_i^* \geq c_i) = E(y_i) = \Phi(\alpha + \beta c_i)$, where (by a slight change with respect to our previous notation) $\alpha \equiv \mu / \sigma$ and $\beta \equiv -1/\sigma$. The median can be written as $\mu = -\alpha / \beta$ and estimated by the method of maximum likelihood as $\hat{\mu} = -\hat{\alpha} / \hat{\beta}$.[4]

---

[1]We will follow the standard statistical practice (Silvey, 1980) of writing out the expressions for the experimental designs for non-linear functions in terms of the model's parameters. In general, some or all of those parameters will be unknown. The implications of unknown parameters in the expressions for the thresholds are taken up in Section 3.7.

[2]Variation in the thresholds is *essential* to estimate μ and σ separately. Suppose that a single value of the threshold is used for all the respondents. In that case the single threshold becomes absorbed into the constant term of the discrete choice model. The intercept becomes $(\mu\text{-c})/\sigma$: it may be seen from this expression that only if either μ or σ are known to the researcher prior to the investigation, is it possible to uniquely disentangle the other parameter from the estimate for the constant term. If the design involves only a single threshold value, the information matrix (3) becomes singular.

[3]In a binary discrete choice contingent valuation survey the respondent is asked whether or not he would pay a specific amount for a proposed change in the provision of a public good. The proposed amount, which is varied across the respondents, is the threshold in the discrete choice model, and the population's willingness to pay for the good in question, which is of primary interest to the researcher, is the latent variable. This eliciting procedure is also sometimes referred to as the "closed-ended", "take-it-or-leave-it", or "referendum" method.

[4]The median is the quantile corresponding to p=0.5; for symmetric distributions like the normal the median is equal to the mean of the population.

There are several objectives that one may want to pursue with respect to the choice of the thresholds. One of the most obvious is to choose them so as to directly minimize the asymptotic variance around our parameter of interest. This objective defines c-optimal designs (Silvey, 1980). In general such designs vary with the statistic of interest. As an intuitive example, one may think that designs that are supposed to produce good estimates for the center of a distribution are likely to concentrate the thresholds around that center, whereas the thresholds are more likely to be spread out if the goal is to estimate the variance efficiently.

The asymptotic variance may be replaced with other related criteria. For instance, instead of minimizing the length of the standard confidence interval about the median as follows from the direct variance minimization approach, one may choose to minimize the length of the *fiducial* interval around the median, a related concept frequently used in the biometrics literature, particularly when dealing with the ratio of two parameters (Finney, 1978).[5]

Another alternative is that the researcher seeks an optimality criterion that is unattached to any particular value of p or statistic of interest g. For example, if he wishes to consider a design that attempts to estimate both the median μ *and* the scale parameter σ efficiently, a design criterion known as d-optimality is often used. D-optimal designs minimize the confidence ellipsoid around θ for a given confidence level. The confidence ellipsoid is defined by the set of values of θ such that $(\theta-\hat{\theta})' I(\theta)(\theta-\hat{\theta})$ is less than a constant value that depends on the confidence level. Since the volume of the ellipsoid is proportional to $[|I(\theta)|]^{-1/2}$, the criterion reduces to the maximization of the determinant (or the log determinant) of the Fisher information matrix.

[5]Fiducial inference, proposed by Fisher (1956), is interpreted by some as a sort of a compromise between classical and Bayesian inference. In general, fiducial inference allows for uncertainty about the parameter by supposing that the parameters are unknown and that the only information available about them is the sample draw. Given, for instance, a single draw x from a normal population with mean μ and standard deviation 1, then x is considered fixed and μ is assumed distributed as a normal random variable with mean μ and variance 1, and $x \pm 1.96$, the fiducial limits, include all but 5% of the distribution. Fiducial and classical confidence limits coincide in a number of cases, the most notable being that of the mean of a normal distribution with unknown variance σ^2, given a sample of observations on a continuous scale.

All of the design criteria presented above assign thresholds to the experimental subjects to optimize the value of some functional defined on the information matrix.[6] It is also possible to pursue designs without explicit reference to an optimality criterion. Experimental designs based on dividing the respondents (usually uniformly) among a preselected, larger number of thresholds (usually regularly spaced), such as the ones we consider in Section 3.4, are intuitively appealing. In practice the latter designs are frequently preferred to the designs that optimize a specific objective function on the grounds of their comparable or better performance over a wider range of quantiles (Section 4.2), and when there is some uncertainty about the distribution of the latent variable (Section 3.6).

In what follows, we assume, as is common in surveys or experiments, that the total sample size n is predetermined and given to the survey or experiment designer. Another element of the design that is often predetermined is k, the total number of distinct thresholds allowed. We will further assume that an equal number of sample units are randomly assigned among the thresholds. The assumption of an equal division among the thresholds is not restrictive if the distribution is symmetric, the statistic of interest is the median, and there are an even number of thresholds.[7] Otherwise, optimal designs may require a non-equal division of the observations among the thresholds.[8]

---

[6] In addition to c-optimal and d-optimal designs, and designs that minimize the length of the fiducial interval around the median, a number of design criteria also defined on the information matrix are available, such as e-optimal and g-optimal designs (Silvey, 1980). These criteria for optimality involve statistics of interest that are linear combinations $a'\theta$ of the parameter vector θ. For fixed vector a, the variance of the linear combination is $a'I^{-1}(\theta)a$. If a number of alternative a's are considered over the design space, the design that minimizes the maximum of $a'I^{-1}(\theta)a$ over all a's is the g-optimal design. The c-optimality criterion is a special case of the g-optimality criterion. Another special case of g-optimality is the e-optimal design, which imposes the additional constraint that $a'a = 1$ and is equivalent to maximizing the minimum eigenvalue of $I(\theta)$.

[7] The framework developed in this paper readily extends to the log-normal distribution, frequently assumed for many economic variables. The c-optimal, fiducial and d-optimal designs will consist of two design points, each with an equal number of observations, placed symmetrically around the median of the log of the distribution but asymmetrically around the actual median.

[8] Fiducial designs for percentiles other than the median will generally require an unequal division of sample units between two thresholds placed asymmetrically around the percentile of interest. This division of sampling units between the two design points becomes more unequal as the percentile of interest moves away from the median. Duffield and Patterson (1991) consider a problem where the sample size is assumed to be given and

As a measure of the efficiency loss when moving to situations in which only the indicator variable y_i can be observed, we will use an operative version of the Pitman relative efficiency (Lehmann, 1983). The specific relative efficiency measure we use is the ratio of the asymptotic variance of the MLE for $g = \mu$ from the uncensored data to the asymptotic variance of the probit MLE for the same g:

(5) $$R.E. \equiv \frac{\mathrm{Var}(\mu | \hat{y}_i^*)}{\mathrm{Var}(\mu | \hat{y}_i^*; c_i)}.$$

The reciprocal of the relative efficiency provides an index which has an immediate interpretation in terms of how much larger, in percentage terms, the sample size should be with binary indicator observations relative to the size required to achieve the same level of efficiency with observations on y_i^*.

3.1 C-optimal designs.

The c-optimality approach is structured to find the design that directly minimizes the variance around the statistic of interest. We can exploit the parametric nature of our problem of estimating the median of a normal distribution to estimate g by the method of maximum likelihood and find the design that minimizes the variance of $\hat{g}$. To do so, we must first determine the distribution of $\hat{g}$ and find the expression for the variance of $\hat{g}$. The large sample properties of the MLE may be invoked to show that under suitable regularity assumptions[9] $(\hat{\alpha} - \alpha)$ and $(\hat{\beta} - \beta)$ converge in distribution at rate $\sqrt{n}$ to a bivariate normal variable centered around zero and with covariance matrix equal to the inverse asymptotic Fisher information matrix. The distribution of $g(\hat{\alpha}, \hat{\beta})$ may be derived after approximating $\hat{g}$ with a Taylor series

---

the number of thresholds to be used and their values are also predetermined. In this situation, one will also generally get an unequal division of sample units to the thresholds, even for the median of a symmetric distribution, unless by chance the predetermined thresholds are symmetrically distributed around the median.

[9]See Amemiya (1985). Among the other conditions, the asymptotic information matrix is required to be finite and nonsingular.

expansion of $\hat{g}$ around the true value of the parameters. Typically such a series is truncated at a finite order, say r. If g is continuously differentiable on the parameter space and $\partial g / \partial \theta$ has always full rank, a first-order approximation is deemed sufficient, all higher-order terms being negligible for large sample size (White, 1984). In that case, the normality result for $g(\hat{\alpha},\hat{\beta})$ and the asymptotic variance given in Section 2.1 hold.

To derive the c-optimal design for $\hat{g}$ we use a standard theorem from the optimal design literature, Caratheodory's theorem (see Silvey, 1980, or Chernoff, 1979), which establishes that the c-optimal design for $\hat{g}$ requires *no* more than two design points. This finding agrees with the intuitive notion that only two points are needed to fit a straight line with intercept α and slope β, which is the argument of the standard normal cdf $\Phi(\bullet)$ in our binary choice model. We denote those thresholds by c_1 and c_2.[10] Because of the symmetry of the normal distribution and the fact that the quantile of interest is the median, without loss of generality those thresholds are constructed to be symmetric about the median of the population (Finney, 1978). Let δ be the distance between each threshold and the median: $c_1 = \mu_0 + \delta$, and $c_2 = \mu_0 - \delta$, where μ_0 is the suspected median of the underlying distribution. With a slight change of notation we now rewrite the log likelihood function for this problem as $m_1 \log \Phi(\theta_1) + m_2 \log \Phi(\theta_2)$, with Φ the cdf of the standard normal distribution, $\theta_1 = \beta[\mu_0 - \mu + \delta]$ and $\theta_2 = \beta[\mu_0 - \mu - \delta]$, β being the negative reciprocal of the scale parameter σ, and m_1 and m_2 being the counts of the responses at c_1 and c_2. Provided that neither observed response rate m_1 / n_1 or m_2 / n_2 is zero or one, the MLE's $\hat{\theta}_1, \hat{\theta}_2$ are well-behaved and asymptotically normal (see Amemiya, 1985), even if the two thresholds should coincide. They are also asymptotically independent, as is easily shown by working out the information matrix for θ_1 and θ_2.

After some algebraic manipulation, μ can be shown to be equal to $2\delta\theta_1 / (\theta_1 - \theta_2) - c_1$. The median is thus estimated by the method of maximum likelihood as $2\delta\hat{\theta}_1 / (\hat{\theta}_1 - \hat{\theta}_2) - c_1$. A

[10]The optimal design may in some instances turn out to be a one-point design. Our notation is adequate for those situations, since a one-point design is obtained for $c_1 \equiv c_2$.

Taylor series expansion of the *r*th order around the true θ_1 and θ_2 can be taken to approximate $2\delta\hat{\theta}_1/(\hat{\theta}_1-\hat{\theta}_2)$ and used to develop an approximation to the distribution of $\sqrt{n}(\hat{\mu}-\mu)$. In what follows we denote the approximation based on the *r*th order expansion with $\sqrt{n}(\hat{\mu}-\mu)_{[r]}$.

It will be easiest to look at the asymptotic mean square error (MSE) of the maximum likelihood estimate for μ. Because the asymptotic MSE is the sum of the squared asymptotic bias and the asymptotic variance, effectively a design that minimizes the MSE reduces to a c-optimal design if the estimator is asymptotically unbiased, which we show is the case below. Applying the properties of asymptotic normality and the asymptotic independence of $\sqrt{n}(\hat{\theta}_1-\theta_1)$ and $\sqrt{n}(\hat{\theta}_2-\theta_2)$, we derive the asymptotic mean square error of the estimate for μ based on the *r*th order expansion ($MSE_{[r]}$):

$$\mathrm{MSE}_{[r]} \equiv \mathrm{E}\left\{\left[\sqrt{n}(\hat{\mu}-\mu)\right]^2_{[r]}\right\} = 4\delta^2\left\{\sum_{j=1}^{r}\frac{\theta_2^2}{(\theta_1-\theta_2)^{2(j+1)}}\frac{\mathrm{E}\left[\sqrt{n}(\hat{\theta}_1-\theta_1)\right]^{2j}}{n^{j-1}}+2\sum_{i=1}^{r-1}\sum_{j=1}^{\mathrm{int}[\frac{r-i}{2}]}\frac{\theta_2^2}{(\theta_1-\theta_2)^{2(i+j+1)}}\frac{\mathrm{E}\left[\sqrt{n}(\hat{\theta}_1-\theta_1)\right]^{2(i+j)}}{n^{i+j-1}}\right\}+$$

$$+4\delta^2\left\{2\sum_{i=1}^{r-1}\sum_{j=1}^{\mathrm{int}[\frac{r-i}{2}]}\frac{\theta_1^2}{(\theta_1-\theta_2)^{2(i+j+1)}}\frac{\mathrm{E}\left[\sqrt{n}(\hat{\theta}_2-\theta_2)\right]^{2(i+j)}}{n^{i+j-1}}+\sum_{j=1}^{r}\frac{\theta_1^2}{(\theta_1-\theta_2)^{2(j+1)}}\frac{\mathrm{E}\left[\sqrt{n}(\hat{\theta}_2-\theta_2)\right]^{2j}}{n^{j-1}}\right\}.$$

After taking the expectations and rearranging terms, we obtain, for $r\geq 2$:

$$MSE_{[r]} = 4\delta^2\left\{\frac{\theta_2^2}{(\theta_1-\theta_2)^4}v_1+\frac{\theta_2^2}{(\theta_1-\theta_2)^{2(r+1)}}\frac{\left[2^r\Gamma(r+1/2)/(\pi^{1/2})\right]v_1^r}{n^{r-1}}+3\sum_{k=2}^{r-1}\frac{\theta_2^2}{(\theta_1-\theta_2)^{2(k+1)}}\frac{\left[2^k\Gamma(k+1/2)/(\pi^{1/2})\right]v_1^k}{n^{k-1}}\right\}+$$

$$+4\delta^2\left\{\frac{\theta_1^2}{(\theta_1-\theta_2)^4}v_2+\frac{\theta_1^2}{(\theta_1-\theta_2)^{2(r+1)}}\frac{\left[2^r\Gamma(r+1/2)/(\pi^{1/2})\right]v_2^r}{n^{r-1}}+3\sum_{k=2}^{r-1}\frac{\theta_1^2}{(\theta_1-\theta_2)^{2(k+1)}}\frac{\left[2^k\Gamma(k+1/2)/(\pi^{1/2})\right]v_2^k}{n^{k-1}}\right\}.$$

and, for $r=1$:

$$MSE_{[1]} = 4\delta^2\left\{\frac{\theta_2^2}{(\theta_1-\theta_2)^4}v_1+\frac{\theta_1^2}{(\theta_1-\theta_2)^4}v_2\right\}.$$

In the expressions for the MSE, v_1 is the asymptotic variance of $\sqrt{n}(\hat{\theta}_1-\theta_1)$ and is given by $1/[\lambda w(\theta_1)] = \Phi(\theta_1)[1-\Phi(\theta_1)]/[\lambda\phi^2(\theta_1)]$, λ is the fraction of the sample size assigned to the first threshold, v_2 is the asymptotic variance of $\sqrt{n}(\hat{\theta}_2 - \theta_2)$ and is given by $1/[(1-\lambda)w(\theta_2)] = \Phi(\theta_2)[1-\Phi(\theta_2)]/[(1-\lambda)\phi^2(\theta_2)]$, $\Gamma(\bullet)$ is the gamma function, and Int($\bullet$) indicates the integer part of the expression in parenthesis.[11]

Note that the terms of order higher than one are negligible for large sample size, *provided* that the two thresholds are far apart, since powers of $(\theta_1 - \theta_2) = 2\beta\delta$ (the distance between the two design points) appear at the denominator of each term in the MSE. However, as the thresholds are drawn closer and closer together until they virtually coincide, the first-order expansion fails to provide an adequate approximation to $\sqrt{n}(\hat{\mu} - \mu)$, reflecting the fact that the vector $\left(\frac{\partial\mu}{\partial\theta_1} \;\; \frac{\partial\mu}{\partial\theta_2}\right)'$ will not have full rank as required for the asymptotic normality and consistency of $\hat{\mu}$. We prove the following proposition:

Proposition 1. The c-optimal design for μ (if it exists) is a two-point design.

Proof: see Appendix A. (The proof essentially works by using Caratheodory's theorem to show that the c-optimal design is either a one- or a two-point design, and then rules out the one-point design.[12])

---

[11]The asymptotic bias of the estimate for μ may be shown to be, for order of expansion r:

$$2\delta \sum_{j=1}^{\text{int}(r/2)} \left\{ \frac{\theta_2}{(\theta_1 - \theta_2)^{2j+1}} \frac{2^j \Gamma(j+1/2)/(\pi^{1/2})}{n^{\frac{2j-1}{2}}} v_1^j + \frac{\theta_1}{(\theta_1 - \theta_2)^{2j+1}} \frac{2^j \Gamma(j+1/2)/(\pi^{1/2})}{n^{\frac{2j-1}{2}}} v_2^j \right\},$$

which is zero for any fixed θ_1 not equal to θ_2. As a result, the MSE-minimizing design is equivalent to the c-optimal variance-minimizing design.

[12]Caratheodory's theorem states that each point s^* in the convex hull S^* of any subset S of the n-dimensional space can be represented in the form $s^* = \sum_{i=1}^{n+1} \alpha_i s_i$, where $\alpha_i \geq 0$, $\sum_{i=1}^{n+1} \alpha_i = 1$, and $s_i \in S$. If s^* is a boundary point of the set S^*, then α_{n+1} can be set equal to zero. Chernoff explains some implications of Caratheodory's theorem for information matrices (theorem 2.1.2, p. 66): because of the symmetry of information matrices, if l is the number of unknown parameters, the information matrix can be completely described by a vector of $l(l+1)/2$ elements, and for any experimental design with number of design points exceeding $[l(l+1)/2+1]$ it is always possible to find a design that produces an identical information matrix but has the number of points less than or equal to $[l(l+1)/2+1]$. In our problem Caratheodory's theorem implies that the optimal design requires no more than two design points.

The c-optimal design point, if it exists, is thus a two-point design ($\delta > 0$). One additional result that follows from the proof of the proposition (see Appendix A) is that the situation with $\delta = 0$ (a one-point design) is an instance in which the approximation to the asymptotic distribution of $\sqrt{n}(\hat{\mu}-\mu)$ based on the first-order expansion is inadequate.

Confirming the proposition, at r=1 the optimization program for minimum MSE essentially fails because, while the distance between the thresholds tends to zero, the objective function rapidly increases towards infinity. In light of this problem, we examine the issue of how large r has to be for the approximation of $\sqrt{n}(\hat{\mu}-\mu)$ to be adequate when we look for the values of θ_1 and θ_2 that minimize the asymptotic mean square error. We expected that, if the rth-order approximation is adequate, the optimal design would not change much as higher-order terms are added, whereas there would be an appreciable change in the optimal design using an approximation of order r if indeed terms of higher order are not negligible. We found the optimal design points for increasing values of the order of the expansion, r, over a wide range of commonly used sample sizes (500, 900, 1,000, 5,000, and 10,000).[13] For r=2 there is a well-behaved and non-degenerate solution consisting of two distinct points. Going to r=3, the two optimal design points are considerably farther away from the median, usually over 30% farther apart. For r>3 the thresholds are placed even further apart; however, the increase in the distance between the points tends to be quite small relative to that observed when moving from r=2 to r=3. For instance, with n=500, the difference between the optimal δ's at r=3 and at r=16 is 3%, for n=1000 the difference is 2%, for n=10000 it is less than 1%.

Thus it is clear that given our objective function the approximation to $r \geq 3$ is always justified, and one may want to consider approximations to the fourth or higher orders in small

[13]Once the optimal θ_1 and θ_2 are found, the distance δ between the thresholds and the median would be derived from the optimal θ using the definitions $\theta_1 = \beta[(\mu_0 - \mu) + \delta]$, $\theta_2 = \beta[(\mu_0 - \mu) - \delta]$, the value of the true median μ and that of the assumed median μ_0. It is obvious that the researcher can do no better than treating the assumed median, μ_0, as equal to the true median, μ. This assumption yields $\theta_1 = \beta\delta, \theta_2 = -\beta\delta$, and the actual thresholds $c_1 = \mu + \delta$ and $c_2 = \mu - \delta$.

sample sizes.[14] As n increases, the distance between the thresholds and the median becomes smaller.[15]

Our two-threshold design differs from the single-threshold design often discussed in the biometrics literature (*e.g.*, Wu, 1988), which implicitly assumes that the information matrix is always nonsingular and the terms of order higher than one are always negligible. There the variance minimizing design assigns a single value of the threshold, equal to the median itself, to all respondents. If, however, we look at a somewhat different design problem where the scale parameter σ of the normal distribution is known and does *not* have to be estimated, the design that minimizes the asymptotic variance around the median *is* indeed the one that concentrates all the thresholds at the median. The theoretical (*i.e.*, in correspondence to the true values of the parameters) relative efficiency of the binary choice estimate for μ produced by this design is $2/\pi$ or 0.6366. Such relative efficiency requires that even under the most ideal conditions in order to preserve the efficiency of the estimate for the median the sample size must be increased by over 150% when moving from a continuous choice to a binary choice situation .[16]

---

[14]The deviations between the distances from the median with a third-order and higher-order expansions can be substantially larger for sample size smaller than 100. For instance, for n=50, the difference between the indicated thresholds for r=3 and r=16 was about 13%.

[15]For example, for n=500 the actual thresholds are placed at $\pm0.3333\sigma$ from the median for r=1, at $\pm0.4374\sigma+\mu$ for r=3, at $\pm0.4434\sigma+\mu$ for r=4, at $\pm0.4492\sigma+\mu$ for r=5, at $\pm0.4497\sigma+\mu$ for r=6, at ±0.4502 for r=7 and r=8, at $\pm0.4503\sigma+\mu$ for r=9 and r=10, and at $\pm0.4504\sigma+\mu$ for r greater than 10. For n=900, the placement of the thresholds is at $\pm0.2887\sigma+\mu$ for r=1, jumps to $\pm0.3792\sigma+\mu$ for r=2 and quickly converges to $\pm0.3873\sigma+\mu$ (ascending from r=7). For n=1000 the optimal thresholds are $\pm0.2813\sigma+\mu$ for r=2, $\pm0.3695\sigma+\mu$ for r=3 and $\pm 0.3770\sigma+\mu$ for r higher than 5. For n=5000 the optimal design starts at $\pm0.1890\sigma+\mu$ for r=2, moves to $\pm0.2486\sigma+\mu$ with r=3 and quickly attains the values of $\pm0.2508\sigma+\mu$ for r greater than 4. For a sample size of 10,000 the optimal design points are $\pm0.1591\sigma+\mu$ for r=2, $\pm0.2093\sigma+\mu$ for r=3, $\pm0.2100\sigma+\mu$ for r=4 and $\pm0.2106\sigma+\mu$ from r=5 and up. (A graph of these design points is available from the Authors.)

[16]The theoretical relative efficiency for the median attained using the variance-minimizing design is 0.75 for the logistic distribution, which originates the binary logit model, 0.48 for each of largest and smallest extreme value distributions, which give rise to the log-log and the complementary log-log link functions, 0.81 for the Cauchy distribution and 1.00 for the double exponential. With the double exponential distribution, no loss of efficiency is incurred when moving to a binary data sample because the MLE for the median from uncensored observations is the sample median (Alberini and Carson, 1991).

3.2 Fiducial designs.

Closely related to c-optimality is the design method based on minimizing the length of the fiducial interval around the median for a given fiducial level. Fiducial designs have a long history in the biometrics literature, going back to R.A. Fisher (1956). Much of the popularity of fiducial designs is due to Fieller's theorem, which allows one to obtain an expression for the fiducial limits for the ratio of the mean values of two random variables having a bivariate normal distribution. Finney (1978) provides an example of its use in deriving the fiducial interval around the median $\mu = -\alpha / \beta$ for a given fiducial level. Minimizing the length of the fiducial interval with respect to the thresholds is equivalent to minimizing the half-length of the fiducial interval or its square:

$$\frac{t^2\sigma^2}{(1-b)^2}\left\{\frac{Var(\hat{\mu})}{\sigma^2} - t^2\sigma^2\left[Var(\hat{\alpha})Var(\hat{\beta}) - Cov^2(\hat{\alpha},\hat{\beta})\right]\right\},$$

where $b \equiv \dfrac{t^2 Var(\hat{\beta})}{\beta^2}$ and t is a standard normal deviate. In that expression, b can be rewritten as $b \equiv \dfrac{t^2}{\sum n_j w_j (z_j - \bar{z})^2}$, $\dfrac{Var(\hat{\mu})}{\sigma^2}$ as $\dfrac{\sum n_j w_j z_j^2}{\left[\sum n_j w_j\right]\left[\sum n_j w_j (z_j - \bar{z})^2\right]}$, and $Var(\hat{\alpha})Var(\hat{\beta}) - Cov^2(\hat{\alpha},\hat{\beta})$ as $\dfrac{1}{\sigma^2\left[\sum n_j w_j\right]\left[\sum n_j w_j (z_j - \bar{z})^2\right]}$, with $z_j \equiv \alpha + \beta c_j$, $\bar{z} \equiv \dfrac{\sum n_j w_j z_j}{\sum n_j w_j}$, $w_j \equiv \dfrac{\phi^2(z_j)}{\Phi(z_j)[1-\Phi(z_j)]}$, and all the summations are from $j=1$ to k. The square of the half-length of the fiducial interval can be equivalently expressed, neglecting the scale factor σ^2, as

(6) $$\frac{t^2}{(1-b)^2}\left[\frac{1-b}{\sum n_j w_j}+\frac{\bar{z}^2}{\sum n_j w_j (z_j-\bar{z})^2}\right],$$

which is a function of the weighted average $\bar{z}$ with weights w_j and the estimated variance from the weighted regression of the z_j's on a constant term. The requirement that the design be symmetric about the median implies that $\bar{z} \equiv 0$, while equal division of the sample units among the thresholds implies that $n_j \equiv \frac{n}{k}$ for all j's. The objective function (6) thus simplifies to:

(7) $$\frac{t^2\left[\sum w_j z_j^2\right]}{\left[\frac{n}{k}\sum w_j z_j^2 - t^2\right]\left[\sum w_j\right]},$$

if b is less than 1, which is required to obtain finite fiducial intervals (Abdelbasit and Plackett, 1983). Note that with the fiducial method, the optimal thresholds and the efficiencies depend on the sample size n.[17]

3.3 D-optimal designs.

D-optimal designs maximize the value of the log determinant of the Fisher information matrix (3). Neglecting the scale factor σ^2, the expression to be maximized is the log of

(8) $$\left[\sum_{i=1}^{n} w_i z_i^2\right]\left[\sum_{i=1}^{n} w_i\right]-\left[\sum_{i=1}^{n} w_i z_i\right]^2.$$

---

[17]Each optimal threshold satisfies the first order condition:

$$\frac{d(w_j z_j^2)}{dz_j}\left[\frac{n}{k}\sum w_i z_i^2\right]\left[\sum w_j\right]-\left[\sum w_j z_j^2\right]\left[\frac{n}{k}\frac{d(w_j z_j^2)}{dz_j}\left[\sum w_j\right]+\frac{dw_j}{dz_j}\left[\frac{n}{k}\sum w_j z_j^2 - t^2\right]\right]=0$$

where $\frac{dw_j}{dz_j} = -2z_j w_j - w_j^2\frac{1-2\Phi(z_j)}{\phi(z_j)}$ and $\frac{d(w_j z_j^2)}{dz_j}=\frac{dw_j}{dz_j}z_j^2+2w_j z_j$.

For symmetric distributions, the d-optimal design points are placed symmetrically about the median. With normal y^*, unconstrained maximization of (8) or its log yields a design which, for n even, places $n/2$ observations at $z = \pm 1.138101$ (Ford, Torsney and Wu, 1992). If the restriction of equal division of the sample units among a predetermined number of thresholds k is imposed, for k even, the solution to the constrained maximum problem coincides with the solution of the unconstrained maximum problem (Abdelbasit and Plackett, 1983). For odd k, the constrained d-optimal designs collapse to three-point designs (the median itself being one such design point). Such simplifications make for easier data collection planning and are unsurprising since only two points are necessary to estimate the two parameters that characterize the normal distribution[18].

3.4 The "percentiles" method.

All of the designs considered thus far explicitly optimize some functional defined on the information matrix for the parameters of the model, and require the researcher to make assumptions on the values of those parameters prior to finding the solutions to the optimization problems. If the assumed values for the parameters do not coincide with the true values, the actual thresholds of the two-point designs may turn out to be quite far out in the tails or in only one tail of the distribution, where they would adversely affect the efficiency of the final estimate for μ.

The designs that we consider in this section are comprised of more than two, regularly spaced, thresholds, and have been derived following the notion that increasing the number of thresholds and spreading them more widely around the suspected median of the underlying distribution may increase the likelihood that at least one of the thresholds used is sufficiently

[18]The same properties are shown to hold for the fiducial designs (see Abdelbasit and Plackett, 1983) and for the c-optimal designs of Section 3.1. Each d-optimal threshold satisfies the first order condition,

$$\frac{dw_j}{dz_j}\left[\sum w_j z_j^2\right]+\frac{d\left(w_j z_j^2\right)}{dz_j}\left[\sum w_j\right]-2\left[\sum w_j z_j\right]\frac{d\left(w_j z_j\right)}{dz_j}=0.$$

close to the true median. These designs, which we term ***percentile-based*** designs, are ***ad hoc,*** but intuitively appealing and allow one to trace out the empirical distribution of the latent variable.

The percentile method guarantees a fixed number of thresholds, *k*, to be chosen by the researcher. The thresholds are then found as the points that cumulate $\left(\frac{j}{k+1}\right)100\%$ of the normal distribution, for *j*=1, 2,...,*k* (so that for *j*=2, ...,*k*, $\Phi(z_j)-\Phi(z_{j-1})=\frac{1}{(k+1)}$ and $\Phi(z_1)=\frac{1}{(k+1)}$, where, as before, c_j is the *j*th actual threshold, and $z_j=\alpha+\beta c_j$): the area under the probability density function of the latent variable between any two adjacent thresholds is constant. For instance, if *k*=4, the design points are the quintiles of the underlying normal distribution, for *k*=3 the design points are the quartiles, for *k*=9, the design points are the deciles, etc. The respondents are evenly divided among the thresholds thus found, as is compatible with the equi-probability requirement imposed to fix the thresholds. The design is symmetric around the median for any *k*.

3.5 Covariates.

Econometricians are usually interested in situations in which covariates are present, and the coefficients on these covariates are often the primary focus of the study. While the issue of optimal choice of the thresholds has been explored for the simple probit model with two parameters (α and β), one control variable (the threshold), and no other regressors (Chernoff, 1979; Ford, Torsney and Wu, 1992; and Wu, 1988), to our knowledge no solutions to these problems have been explicitly worked out for situations in which explanatory variables are present. In this section, we discuss possible extensions of the case with no covariates to simple situations in which there are one or more covariates.

When explanatory variables are present, the key distinction we will be interested in is whether the individual values of the covariates are known to the researcher in advance of the choice of the thresholds to be used in the study. With respect to the knowledge of the individual x_i^*'s prior to the designation of the thresholds, we can consider two states of the world: (a) one in which the individual x_i^*'s are known in advance of the choice of the thresholds, in which case the optimal values of the thresholds will in general depend on the values of the covariates and the parameters, or, equivalently, on the conditional first moments of the distributions of the y_i^*'s and (b) one in which the individual x_i^*'s are *not* known in advance, or *cannot* be used to create the thresholds.

The first situation (a) can be seen as the case in which the researcher is able to design the study conditionally on the regressors. This situation occurs when the values of the regressors are known prior to conducting the experiment or when they become known to the researcher and available for design purposes during the course of the experiment. With computer-assisted telephone interviewing (Frey, 1989), for instance, the values of the regressors reported by the respondents are entered in the computer immediately and could be used to calculate the optimal thresholds for that respondent. In an experimental setting, it may be possible to prescreen subjects for the values of the covariates of interest.

Suppose that the true model is $y_i^* = x_i\gamma + \varepsilon_i$, where x_i is a vector of regressors, the disturbance term ε_i is normal i.i.d. with mean zero, indicating that the conditional distribution of each y_i^* is normal with median $x_i\gamma$ and variance σ^2. Assume that the individual values of the covariate are known at the stage of creating the design for the survey. The difficulty here lies in that one is usually interested in the vector of parameters γ, so that a design that aims at efficiency effectively looks at a covariance matrix of dimension *lxl*, *l* being the number of elements of γ. One possibility is to use the d-optimality approach to design, and thus look for the design that maximizes the log determinant of the information matrix for γ.

With the regressors known at the time the design is prepared, an alternative design possibility might be to concentrate attention on each conditional median $m_i = x_i\gamma$. In many instances, the experiment is conducted in such a way that there is stratification with respect to the regressors and one is interested in the median m_j in the stratum defined by one of the possible sets of values of the regressors. In this case, one may choose to specialize the general c-optimal design method to each stratum, producing an experimental design whereby each stratum receives two thresholds symmetrically placed around m_j. Knowledge of the conditional medians m_j is completely interchangeable with information about the individual values of the regressor and about the values of the parameters γ. The distance between each of these thresholds and the conditional median is the scale parameter σ_ε multiplied by a factor that decreases with the sample size (see Section 3.1). Subjects are to be equally divided between those two thresholds. Another possibility might be to create a statistic that is believed to be representative of the population and apply the variance-minimizing approach with reference to this statistic.

A very simple example of a model with covariates is given by: $y_i^* = \mu + \gamma d_i + \varepsilon_i$, with ε normal, where *d,* the covariate, is a dummy variable that takes on a value of one for half of the subjects in the sample (say, for $i = 1, 2, \ldots, n/2$) and zero for the remainder ($i = n/2+1, \ldots, n$). For the first $n/2$ respondents the conditional median is $\mu+\gamma$, for the remainder the conditional median is μ. Identification of the parameters of the model requires that at least two distinct values of the thresholds are administered to the respondents. Note, however, that if all the respondents of the first stratum receive a common threshold and all those in the other subsample receive another common threshold, the parameters are still not identifiable because the information matrix is singular, even though there are two distinct values of the threshold. To overcome this problem while keeping our objective function(s) manageable, we first considered c-optimal designs that impose the restriction that both strata are evenly divided between two thresholds each and minimize the asymptotic MSE for, in turn, μ (the conditional mean in the second stratum), γ, $\mu+\gamma$ (the conditional mean in the first stratum), and $\mu+\frac{1}{2}\gamma$ (the overall population mean) for various values of r, the order of the Taylor series expansion used to approximate the distribution of the estimates.[19] As an alternative to these c-optimal designs, the fiducial or the d-optimal approach can be used to find the thresholds for the study (see Appendix B).

The most typical situation in empirical practice is that in which the realizations of the regressors (or, alternatively, the conditional moments of the y_i^*'s) are not known to the researcher at the time plans are made for the thresholds, or cannot be used to formulate the design because of practical concerns. In that case the researcher can do no better than using information about the *unconditional* (or population) moments of y^* to formulate the statistical design. To illustrate this situation consider a model similar to that used earlier, with a constant and a dummy variable $y_i^* = \gamma_0 + \gamma_1 d_i + \varepsilon_i$. To keep the discussion simple, assume that σ_ε^2 is known to be equal to one. Although we might know that the median y^* is equal to γ_0 in one stratum of the population and to $\gamma_0 + \gamma_1$ in the other, we shall assume that it is not known at the design stage which stratum an individual belongs to. In this situation a reasonable option for the researcher interested in estimating γ_0 and γ_1 is to choose a design that is based on the population (or unconditional) moments, such as the single-threshold design consisting of $c = E(y^*) = \gamma_0 + \gamma_1 E(d) = \gamma_0 + 0.5\gamma_1$. The relevant information matrix in a situation in which the individual values of the covariates are not known prior to the survey is the expected *ex ante* information matrix. Its general expression for a model with an intercept term and one regressor is:

$$I(\theta) = \sum_{i=1}^{n} E_{x^*}\left\{ \frac{\phi^2(z_i)}{\Phi(z_i)[1-\Phi(z_i)]} \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix} \right\} \equiv nE_{x^*}\left\{ w(z) \begin{bmatrix} 1 & x^* \\ x^* & x^{*2} \end{bmatrix} \right\}$$

[19]We found that the MSE-minimizing designs followed a pattern similar to that described in Section 3.1: the optimization routine failed for r=1 with thresholds identical to each stratum's conditional mean. For r=2 the first stratum received two thresholds that were symmetric about the conditional median $m_1 = \mu + \gamma$, while the second stratum received two coinciding thresholds equal to $m_2 = \mu$. A considerable increase in the distance between the thresholds and m_1 was noted when moving from r=2 to r=3. The placement was a function of the sample size.

where E_{x^*} denotes the expectation with respect to the distribution of the regressor x^* in the population, $z_i \equiv \gamma_0 + \gamma_1 x_i - c$ and $z \equiv \gamma_0 + \gamma_1 x^* - c$, and c is the single threshold value used in the survey. On taking the expectation with respect to the population distribution of the d's (a binomial with parameter q=0.5), the *ex ante* information matrix for our model is:

$$I^*(\gamma_0,\gamma_1) = nE_d\left\{w(z)\begin{bmatrix}1 & d\\ d & d^2\end{bmatrix}\right\} = \frac{n}{2}w(z|d_i=1)\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} + \frac{n}{2}w(z|d_i=0)\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} = nw(0.5\gamma_1)\begin{bmatrix}1 & 0.5\\ 0.5 & 0.5\end{bmatrix}$$

where $w(z|d_i=1) = w(\gamma_0+\gamma_1-c) = w(0.5\gamma_1)$ and $w(z|d_i=0) = w(\gamma_0-c) = w(-0.5\gamma_1)$ reflect the fact that $(\gamma_0+\gamma_1 d)$ is equal to $(\gamma_0+\gamma_1)$ if d=1 and to γ_0 if d=0. An additional simplification of the expression for $I^*(\gamma_0,\gamma_1)$ follows from the symmetry of $w(z)$ around zero.[20] We note that the present design attains a 50%-50% overall split of the responses between zeros and ones, but that the split *within* each stratum may be very different from 50%-50%. The suggested design clearly uses the *unconditional* model for the underlying y^* and calculates the threshold as a function of the unconditional first moment of y^*. This can be done either by using the unconditional first moment of y^* directly, or by calculating it as the linear combination of the first moment of the regressor and the parameters of the model. For design purposes, information about the *moments* of the regressor, combined with knowledge of the values of the parameters, offers no advantage over using the unconditional moments of y^* directly.

---

[20]Other designs are possible in this situation: for instance, one could conceive of single-threshold designs with $c=\gamma_0$, or $c=\gamma_0+\gamma_1$. A quick comparison of the *ex ante* information matrices associated with those designs, however, shows that the single threshold design with $c=\gamma_0+0.5\gamma_1$ is *always* better than the other two designs with respect to γ_1, and is superior with respect to γ_0 as well for all the values of γ_1 in the range between approximately -2.5 and 2.5. For those values of γ_1 the difference between the information matrix derived from the design with by $c=\gamma_0+0.5\gamma_1$ and the information matrix for one of the other designs considered is a positive definite matrix. As an alternative, one may want to assign a threshold equal to γ_0 to half of the sample units, and a threshold level of $\gamma_0+\gamma_1$ to the remainder of the sample. This design is inferior to the single-threshold design with $c=\gamma_0+0.5\gamma_1$ for any value ofγ_1 in the usual matrix sense (that is, the difference between the inverse information matrix associated with the design that assigns γ_0 to one half of the sample units, and $\gamma_0+\gamma_1$ to the other half, and the inverse information matrix from the design that assigns the threshold $c=\gamma_0+0.5\gamma_1$ to everyone is positive semidefinite).

One interesting issue is whether using the c-optimal design that is based on the unconditional moments of the underlying y^* brings about a loss of efficiency for the statistic of interest with respect to the c-optimal design that is based on the conditional moments. Intuition suggests that there should be such a loss of efficiency. If, in the model with the dummy regressor earlier described, the researcher had known which stratum an individual belongs to at the time of the design, the c-optimal design would have assigned a different level of the threshold to each stratum: $c_1 = \gamma_0$ and $c_2 = \gamma_0 + \gamma_1$. This design would have produced the following information matrix:

$$I(\gamma_0, \gamma_1) = \frac{n}{2} w(z|d_i = 1)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \frac{n}{2} w(z|d_i = 0)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} =$$

$$= \frac{n}{2} w(0)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \frac{n}{2} w(0)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = nw(0)\begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.5 \end{bmatrix},$$

which follows from $z = \gamma_0 + \gamma_1 d - c_2 = \gamma_0 + \gamma_1 - (\gamma_0 + \gamma_1) = 0$ if $d=1$ and $z = \gamma_0 + \gamma_1 d - c_1 = \gamma_0 - \gamma_0 = 0$ if $d=0$. Clearly, this design would have attained 50%-50% overall and stratum splits of the responses between zeros and ones. That the present design is more informative than that based on the unconditional moments of y^* can be seen by comparing the respective information matrices $I(\gamma_0, \gamma_1)$ and $I^*(\gamma_0, \gamma_1)$: because $w(z)$ has a maximum at zero, $I(\gamma_0, \gamma_1)$ is greater than $I^*(\gamma_0, \gamma_1)$ element by element. The covariance matrix for $\hat{\gamma}_0$ and $\hat{\gamma}_1$ obtained by inverting $I(\gamma_0, \gamma_1)$ is thus *smaller* than that obtained by inverting $I^*(\gamma_0, \gamma_1)$ in the usual matrix sense, confirming the intuitive notion that when the design must be based on the unconditional moments of y^* there is some loss of information relative to a design based on the conditional moments. The extent of the loss of efficiency depends on the value of the parameter γ_1 and is presumably less serious as γ_1 approaches zero.[21]

[21]The loss of efficiency of the design based on the unconditional median happens because, when the statistic of interest is the median, the highest level of information about the center of the distribution is obtained by choosing the value(s) of the threshold that create a 50%-50% split of the responses between zero and one. Both the design based on the conditional medians and that based on the unconditional median aim at an overall

However, even in the situations in which one can do no better than using a design based on the unconditional moments of y^*, after the experiment or survey has been completed and the binary responses have been observed it is still best to *estimate* the model *including* the covariates. To illustrate this claim, we move to another simple model with an intercept term and one covariate: $y_i^* = \gamma_0 + \gamma_1 x_i + \varepsilon_i$, with the regressors i.i.d. and normal with mean zero and variance assumed known and equal to 1. We assume that the parameter of interest is γ_0. If the conditional moments (or, equivalently, the individual values of the covariates) are not known or cannot be used for design purposes, effectively to form the optimal design one refers to the specification $y^* = \gamma_0 + \gamma_1 x + \varepsilon = \mu_{y^*} + \eta$, where $\mu_{y^*} = \gamma_0 + \gamma_1 E(x)$ is the unconditional mean/median of y^*. The new error term, $\eta = \gamma_1 \varepsilon_x + \varepsilon$, where ε_x is the deviation of x from its mean, has mean zero and variance $\gamma_1^2 \sigma_x^2 + \sigma_\varepsilon^2$, which is here equal to $(\gamma_1^2 + 1)$. The unconditional model for y^* reproduces the situation in which there are no covariates, and allows one to derive the c-optimal design by applying the results of Section 3.1 in a straightforward manner: the c-optimal design for γ_0 consists of assigning all the observations to the threshold $c = E(y^*) = \mu_{y^*} = \gamma_0 + \gamma_1 E(x) = \gamma_0$. The expected *ex ante* information matrix is:

$$I^*(\gamma_0, \gamma_1) = \begin{bmatrix} \int w(\gamma_1 x)\phi(x)dx & \int w(\gamma_1 x)x\phi(x)dx \\ \int w(\gamma_1 x)x\phi(x)dx & \int w(\gamma_1 x)x^2\phi(x)dx \end{bmatrix} = \begin{bmatrix} \int w(\gamma_1 x)\phi(x)dx & 0 \\ 0 & \int w(\gamma_1 x)x^2\phi(x)dx \end{bmatrix}$$

where $\phi(x)$ is the standard normal density and $w(\cdot)$ is defined as in Section 3.2. The diagonal elements of $I^*(\gamma_0, \gamma_1)$ depend on the value of the parameter γ_1, a finding that confirms our results for the model with a dummy regressor. For γ_1=1, for instance, this matrix is approximately equal to:

$$I^*(\gamma_0, \gamma_1) \cong \begin{bmatrix} 0.48n & 0 \\ 0 & 0.27n \end{bmatrix},$$

---

frequency of ones in the sample equal to 50%, but while the design based on the conditional medians attains likelihood 50% of a one for *each* individual (or in each stratum of the population), the design based on the unconditional median cannot ensure this.

which yields variances for $\hat{\gamma}_0$ and $\hat{\gamma}_1$ equal to $1/(0.48n)$ and $(1/0.27n)$. If one uses the same design ($c = \gamma_0$) to collect the binary responses and estimates the binary response model without including the covariate, the estimated model only has the intercept term which is equal to the ratio of μ_{y^*} / σ_η. Assuming that the unconditional variance is known, the estimate of the parameter of interest $\gamma_0 = \mu_{y^*}$ can be recovered. Its variance is equal to $\sigma^2_\eta \cdot \frac{\pi}{2n} \cong \frac{3.14}{n}$, which is larger than $1/(0.48n) \cong 2.08/n$ obtained earlier because the former contains a larger residual variance.[22]

3.6 Uncertainty about the parameter values.

In all of the designs considered thus far, the thresholds are functions of the unknown parameters μ and σ of the underlying distribution. For the sake of simplicity we will refer to the model without covariates from here on. We observe that in the Fisher information matrix (3), the argument of each w_i component is $\alpha + \beta c_i$. The objective functions for the c-optimal, the fiducial and the d-optimal designs are functionals defined on the Fisher information matrix,

[22]Finally, we can also calculate the loss of efficiency implied by discrete choice estimation of γ_0 in lieu of estimating this parameter using a sample of exact numerical values for y^*. If the binary response model that omits the covariate (and uses the design based on the unconditional moment $c = \gamma_0$) is compared to a continuous-data model that also omits the covariate, the relative efficiency for γ_0 of the former relative to the latter is 0.6366. If the covariate is included in the continuous-data model, but not in the binary response model, the relative efficiency is approximately 0.3289. When the covariate x is included in both models the relative efficiency for γ_0 is 0.48. The latter relative efficiency is lower than the relative efficiency (0.6366) obtained if the covariate is included in both models, but the binary response data is collected using the design that does use information about the conditional moments of y^*. This design would present each respondent with a different threshold that is formed as a function of the respondent's value of x: $c_i = \gamma_0 + \gamma_1 x_i$. To summarize, when we move from estimating the unconditional model for y^* to estimating the model for y^* that is conditional on the covariate, both the binary response and the continuous data estimates gain efficiency. However, the gain is more pronounced in the continuous data model, which results in a decrease of the relative efficiency of the discrete choice estimates vis-a-vis the continuous data estimates. To recover the maximum attainable relative efficiency (0.6366), it is necessary to collect the binary responses using the c-optimal design based on the conditional moments of the y^*.

which in turn depends on the parameters of the model.[23] If a percentile-based design is preferred, one still needs to know the values of the location and scale parameters to contruct the thresholds. Thus, the actual implementation of the design requires knowledge of σ and μ, the very parameters the researcher wants to estimate.

Three different approaches have been suggested to tackle this problem. The first approach simply replaces the unknown parameters with some initial estimates in the expression for the objective function or the thresholds. This is a one-stage approach in which the initial estimates for the parameters come from a source such as a previous study, a pilot study, or perhaps a pure guess and are treated as the true values of the parameters. The second approach is a sequential one and represents one statistical way of formally treating prior information and updating it. The sequential approach begins with an initial, possibly arbitrary, estimate of the parameters which is improved in the course of *sequential* experiments or surveys. The data collection effort is divided into stages and the estimates from each stage are used for design purposes in the next stage. Each step produces estimates that are more efficient.[24] The third approach is a Bayesian one: the researcher following this approach formally quantifies the level of *a priori* confidence about the true value of the parameters by means of a prior distribution on the parameters (Tsutakawa, 1972, 1980; Kanninen, 1991).

---

[23]From an operative point of view, (8) and (9) can equivalently be optimized with respect to the z_i's. The solutions must be rescaled by the standard deviation and shifted by an amount equal to μ to be on the same scale as the underlying y^*.

[24]The total sample size, number of stages, and rule to assign the number of observations to each stage may be specified prior to the experiment or survey (Abdelbasit and Plackett, 1983; Minkin, 1987). In other cases, the researcher may keep sampling and repeating the sequential procedure without these constraints and bring the experiment or survey to an end when a cost constraint becomes binding (McLeish and Tosh, 1990) or when the precision of the estimate converges to a value that cannot be significantly improved. Using a computer-assisted telephone interviewing (CATI) system it would be possible to sequentially update the parameter estimates with each interview. Kanninen (1993) and Nyquist (1992) follow the sequential approach and present an application to the design of a contingent valuation study in which the variance-minimizing design for μ is adopted at each stage of a sequential design with an updated estimate for μ. A sequential approach of this kind presents the advantage of making the overall design (obtained pooling all the stages) nonsingular while improving the efficiency of the estimates.

In this paper, we deal only with the first approach. In addition to examining the properties of the designs presented when the true values of the parameters are known, we will also analyze by means of a simulation study the behavior of such designs when incorrect initial estimates are used. The evidence from this simulation study hints at how robust the alternative designs are in the presence of biased initial estimates. We hope to deal at length with the sequential and Bayesian approaches in a later paper.

4. Construction of experiment designs

4.1 The c-optimal, fiducial, d-optimal and percentile design points.

Table 1 reports the optimal design points (in standard units) for three different design criteria (the fiducial interval, the d-optimal design and the percentile-based approach) for selected k's. Our optimization routine did not constrain the k optimal thresholds to be all different from each other. We found that for k even and greater than 2, the c-optimal, the fiducial method and the d-optimal designs result in only two distinct threshold values. These values are identical to those of the respective unconstrained designs (Sections 3.1, 3.2 and 3.3). For odd $k > 3$, the constrained designs effectively reduce to three-point designs. We note here that as k odd is allowed to grow towards n, the c-optimal, fiducial and d-optimal designs tend to their unconstrained variants, which, as reported in Sections 3.1, 3.2 and 3.3, consist of two distinct design points. This trend is apparent even at relatively low k's and is demonstrated in Tables 1 for the fiducial and d-optimal designs for k=3, 5 and 9. Moving from k=3 to k=5 shows that fewer observations are placed at the center of the distribution, the remainder being equally divided between two design points that are symmetric around μ and *closer to the center* than the design points for k=3. The same changes are seen when further increasing k from 5 to 9. Although we chose not to report the designs for larger values of k (k odd), it is clear that as k keeps growing, the fraction of the sample assigned to the center tends to zero and the standardized values of the other two design points approach ± 0.37 (for the fiducial design problem) and ± 1.138 (for the d-optimal design problem).

TABLE 1.a

Design Points.

Fiducial Method (95% fiducial level).

k=2, 4, 6, ...	k=3	k=5	k=9
450 ± [0.372581]	300 ± [0.458768]	360 ± [0.417678]	400 ± [0.395715]
	300 [0]	180 [0]	100 [0]

TABLE 1.b

Design Points.

D-Optimal Method.

k=2, 4, 6, ...	k=3	k=5	k=9
450 ± [1.138101]	300 ± [1.298080]	360 ± [1.231471]	400 ± [1.188696]
	300 [0]	180 [0]	100 [0]

TABLE 1.c

Design Points.

Percentiles Method.

k=2	k=3	k=4	k=5	k=6	k=9
450 ± [0.43073]	300 ± [0.67449]	225 ± [0.84162]	180 ± [0.96742]	150 ± [1.06757]	100 ± [1.28155]
	300 [0]	225 ± [0.25335]	180 ± [0.43073]	150 ± [0.56595]	100 ± [0.84162]
			180 [0]	150 ± [0.18001]	100 ± [0.52440]
					100 ± [0.253335]
					100 [0]

Each column in Table 1 provides the number of sample units to assign to the design points (in brackets) for a selected k for a standard normal distribution. The actual thresholds to administer to the respondents are obtained by multiplying these standardized design points by σ and adding μ. Thresholds for the fiducial method are for n=900. For the fiducial method and d-optimality, designs with k even reduce to the respective two-point designs.

The thresholds obtained with the c-optimal and fiducial method are dependent on the sample size n. We use a sample size of 900 because this is roughly the size of many moderate to large surveys and experiments and because 900 is easily divisible by all the values of k in which we are interested. In the cases we considered the c-optimal designs were always very similar to the 95% fiducial designs. For instance, the c-optimal designs for sample sizes equal to 500, 900, 1000, 5000, and 10000 are (in order) $\mu \pm 0.4504\sigma$, $\mu \pm 0.3873\sigma$, $\mu \pm 0.3770\sigma$, $\mu \pm 0.2508\sigma$, $\mu \pm 0.2106\sigma$, respectively. The 95% fiducial interval design points are, following the same order, $\mu \pm 0.4332\sigma$, $\mu \pm 0.3726\sigma$, $\mu \pm 0.3627\sigma$, $\mu \pm 0.2411\sigma$, and $\mu \pm 0.2023\sigma$. Our c-optimal designs are thus virtually identical to the 95% fiducial designs.[25] For this reason we only present the results for the 95% fiducial designs in tabular form.

Since the actual design points must be on the same scale as the underlying y^*, the numbers in Tables 1 must be scaled through multiplying by σ and adding μ. When a design point is zero the corresponding rescaled threshold is the median of y^*. Tables 1 show that the fiducial design points are close to the median while the d-optimal design points are farther away, more toward the tails. This difference in the placement of the design points follows from the fact that a d-optimal design should by definition obtain reasonably efficient estimates for both the scale and the location parameters, while the fiducial design is only trying to obtain an efficient estimate of the latter.

The relative efficiency for the median with respect to a situation in which the individual y_i^*'s are observed with the fiducial method is 0.60519 for $k = 2$ and $n = 900$. The d-optimal designs are always less efficient for the median than the c-optimal and the fiducial designs. The relative efficiency of the d-optimal design is 0.39165 for k=2 or any even k. The highest

[25]We found that the fiducial and the c-optimal designs are identical when the fiducial level is 96.6%. Given the sample size, those designs produce a 96.6% classical confidence interval around μ that has the same length as the 96.6% fiducial interval around μ. The 95% classical confidence interval based on the c-optimal design and the minimum 95% fiducial interval around μ are of almost identical length, the classical interval being only slightly shorter. Further investigation into the properties of the fiducial design showed that, given the sample size, the distance of the fiducial points from the median μ is an increasing function of the desired fiducial level.

relative efficiency afforded by a d-optimal design is achieved with k=3 (0.43657). These results are not surprising: the d-optimal designs maximize the (log) determinant of $I(\theta)$, not relative efficiency. The variance of the estimated μ can, in fact, be *expected* to decrease when we move to the d-optimal design with k=3, because this design places a fraction of the respondents at the median, and clusters the other two design points around the median. The relative efficiency for a percentile-method design is very close to that of the fiducial points for k=2 and k=3 (0.59 and 0.57, respectively). For k equal to 4, the relative efficiency is 0.54, for k equal 6, 0.53, for k equal 8, 0.51. The relative efficiency of the percentile method continues to decline as k is increased.[26]

4.2 Efficiencies at other quantiles.

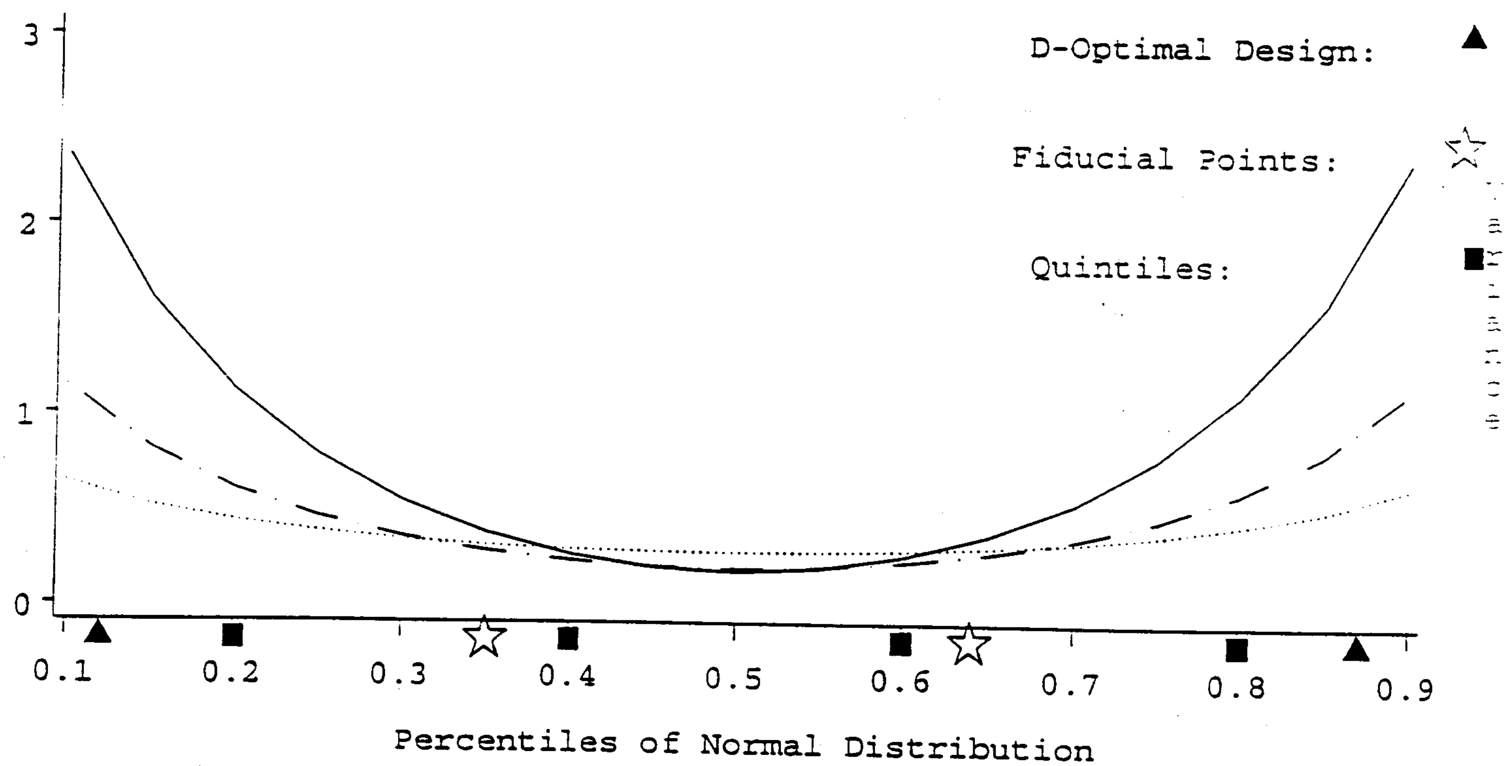
The c-optimal and the fiducial design criteria produce, for any given confidence level, shorter confidence intervals for that quantile than does the d-optimal design, and are thus the best candidates for the design if the researcher's interest lies in one particular quantile of the distribution. However, it is plausible that the sample of data collected using those experimental designs may have to be used to predict other quantiles as well.

Figure 2 displays the efficiencies of the estimated inverse normal cdf $[\Phi^{-1}(p)\hat{\sigma}+\hat{\mu}]$ attained using the design that minimizes the 95% fiducial interval around the median and the d-optimal design, as well as the placement of the respective design points. We have used the unconstrained versions of these designs (k=2). The fiducial design for $p \doteq 0.50$ ensures the most efficient estimates of the median, as expected, and quantiles defined by values of p around

[26]We note here that the close proximity of the c-optimal and fiducial design points to the median tends to result in poor estimates for the variance parameter. The relative efficiency for σ^2 from these designs is approximately 0.04, while with the d-optimal design it is 0.25. The c-optimal design which minimizes the variance about σ^2 by equally splitting the observations between $\pm 1.57\sigma+\mu$ attains a relative efficiency for σ^2 of 0.30. The latter two are considerably farther away from the median than those of the d-optimal design $(\pm 1.14\sigma+\mu)$. For comparison, the percentile designs with k = 4 and k = 6 attain relative efficiencies for σ^2 of 0.10 and 0.11, respectively.

Figure 2.

Variance of Discrete Data Estimator for Percentiles.
Designs are Optimized for Estimating p=0.5 Using a Max. of Four Thresholds.
C-optimal Designs Points are slightly to the Outside of Fiducial Design Points.



Sample Size=900; Mean=10; Standard Deviation=10

Legend: Solid Line: Fiducial Method; - - - : D-optimal Design; -.-.- : Quintiles

0.50, but the d-optimal design does better for quantiles farther away from the median (for p smaller than 0.35 or larger than 0.65). A percentile type of design based on 4 thresholds is included in Figure 2, and is seen to be the best in the approximate ranges of p between 0.35 and 0.45, and between 0.55 and 0.65, suggesting once again that designs specializing with respect to the center of the distribution tend to produce inefficient estimates for other quantiles and are usually outperformed, for those quantiles, by designs whose placement of the thresholds is less concentrated around the center.

4.3 The robustness of the alternative designs.

The experimental designs that we studied (c-optimal, d-optimal, fiducial method, percentiles) can be carried out in practice only if the values of μ and σ are known. Because we consider designs that are completely specified at the beginning of the study, it is necessary to assign values to μ and σ. As a result we decided to explore the impact on the relative efficiency for the median occurring when the values assigned to the parameters at the initial design stage are different from the true values of the parameters.

To assess the impact of assigning biased initial values, we did a simulation study that generated 100 replications of samples of 900 independent observations from a normal random variable with median and standard deviation equal to 10. We then generated the binary indicators using the thresholds corresponding to the c-optimal, d-optimal and fiducial designs (at the 95% fiducial level) for k=2 and odd k's, and the percentile method described in Section 4.3 for k=2, 3, 4, 5, 6, 7, 8, and 9. Our comments are limited to the c-optimal, fiducial and d-optimal designs with k=2 and the percentile designs with k equal to 2, 4, 6, and 8. The percentile designs for odd k have characteristics between those of their adjacent k-even designs, so not much is lost by concentrating attention on the designs for even k.

The first series of replications assumed correct knowledge of the true parameters $\mu = 10$ and $\sigma^2 = 100$; the others introduced initial estimates for the median and the variance that in turn overstate and understate the true parameters by 25%, 50%, and 75%. The median, the variance, and both parameters simultaneously were assigned the various values of the initial estimates; this assignment generates 49 different cases. The values assumed as initial estimates for the median were 2.5, 5, 7.5, 10, 12.5, 15, 17.5, while the initial estimates for the variance were 25, 50, 75, 100, 125, 150, 175.

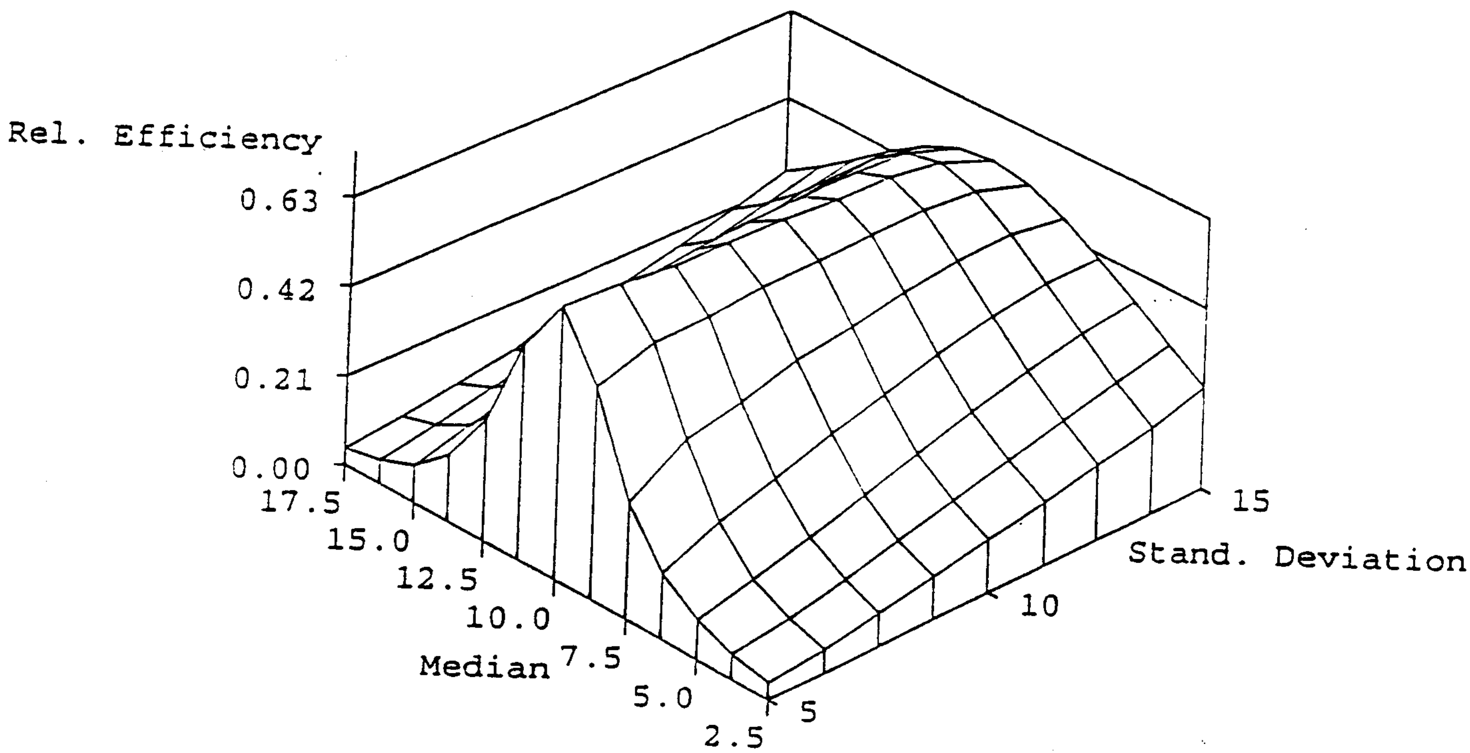
Since we are primarily interested in the relative efficiency, for each replication the estimated variance of $\hat{\mu}$ from the uncensored data was computed. For each k, the appropriate thresholds for each method, given the assumed values for μ and σ, were used to dichotomize y_i^*, then the variance of the probit estimate for μ was computed. The mean relative efficiency for each design was calculated from the 100 replications.

When the initial estimates were equal to the correct values of the parameters, the estimates from the uncensored data, the probit estimates and their relative efficiency were close to the theoretical values for all the design methods. Figures 3 (fiducial), 4 (d-optimal) and 5 (percentile method for k=4) were obtained by graphing the initial values assigned to μ and σ and the resulting average relative efficiencies in a three-dimensional space. It can be seen that each cross-section of the surface (corresponding to a given initial estimate of σ) is symmetric around the true median. The figure for the fiducial design has a sharp peak in relative efficiency at the true value of the median while the d-optimal figure is much flatter and the percentile design for k=4 looks like a flatter version of the fiducial design.

Table 2 displays the mean relative efficiencies for biases of ±25%, ±50%, and ±75% of the true median and variance for the different methods. This table collapses some cells to take account of the symmetry of the effects in over- or understating the initial value of the median, and is a tabular version of the information displayed in the figures. It also contains the average relative efficiencies for the percentile method for k=2, k=4, k=6, and k=8. To make the interpretation of the relative efficiencies easier, the table also reports the sizes of the samples which would have to be drawn in each situation to achieve the same level of efficiency as with a sample of 1000 continuous observations y_i^*.

Figure 3.

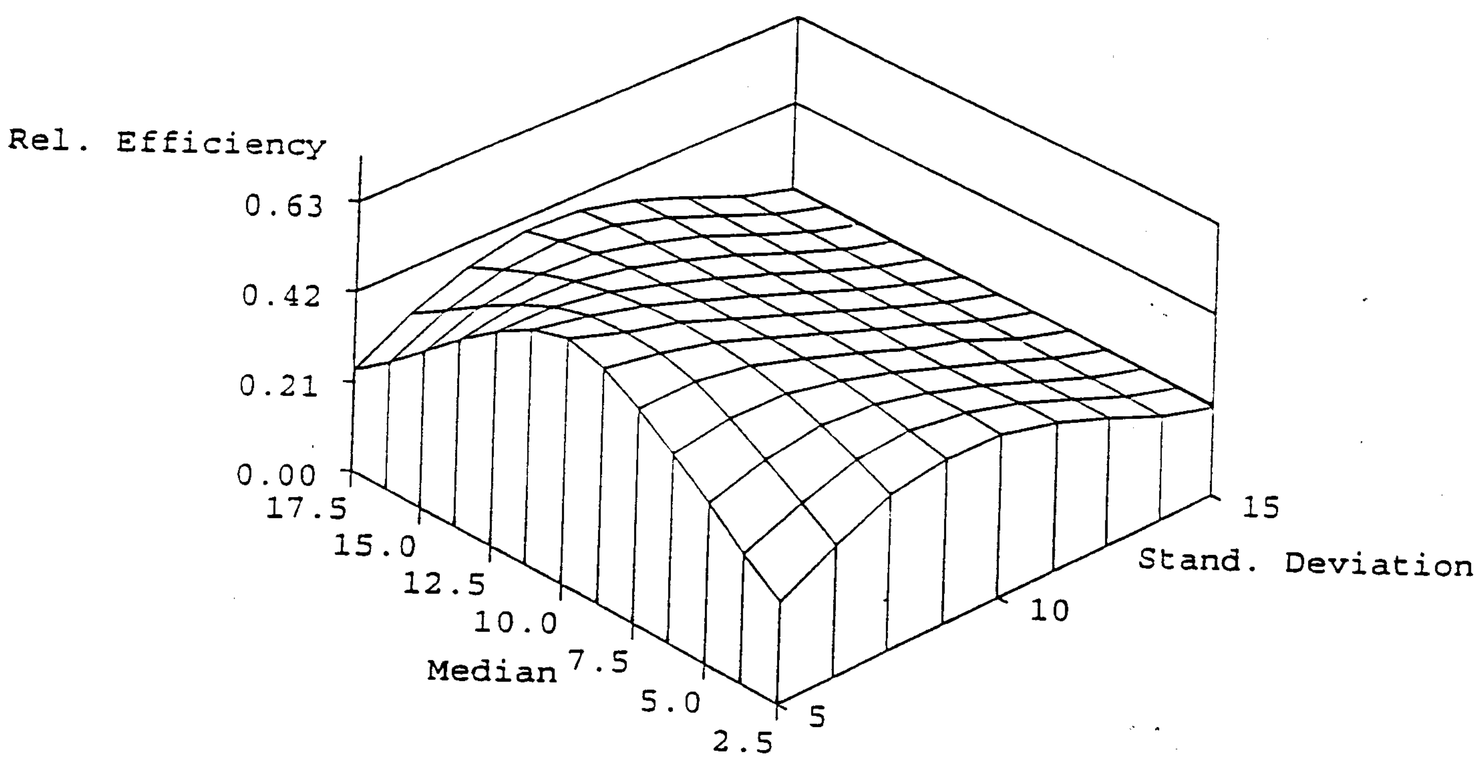
Relative Efficiency of Discrete Data Estimator
Two-point Fiducial Method (95% fiducial level)



Average Relative Efficiencies Over The Replications
As A Function Of The Initial Parameter Estimates
True Value Of Median: 10; True Value Of Standard Deviation: 10

Figure 4.

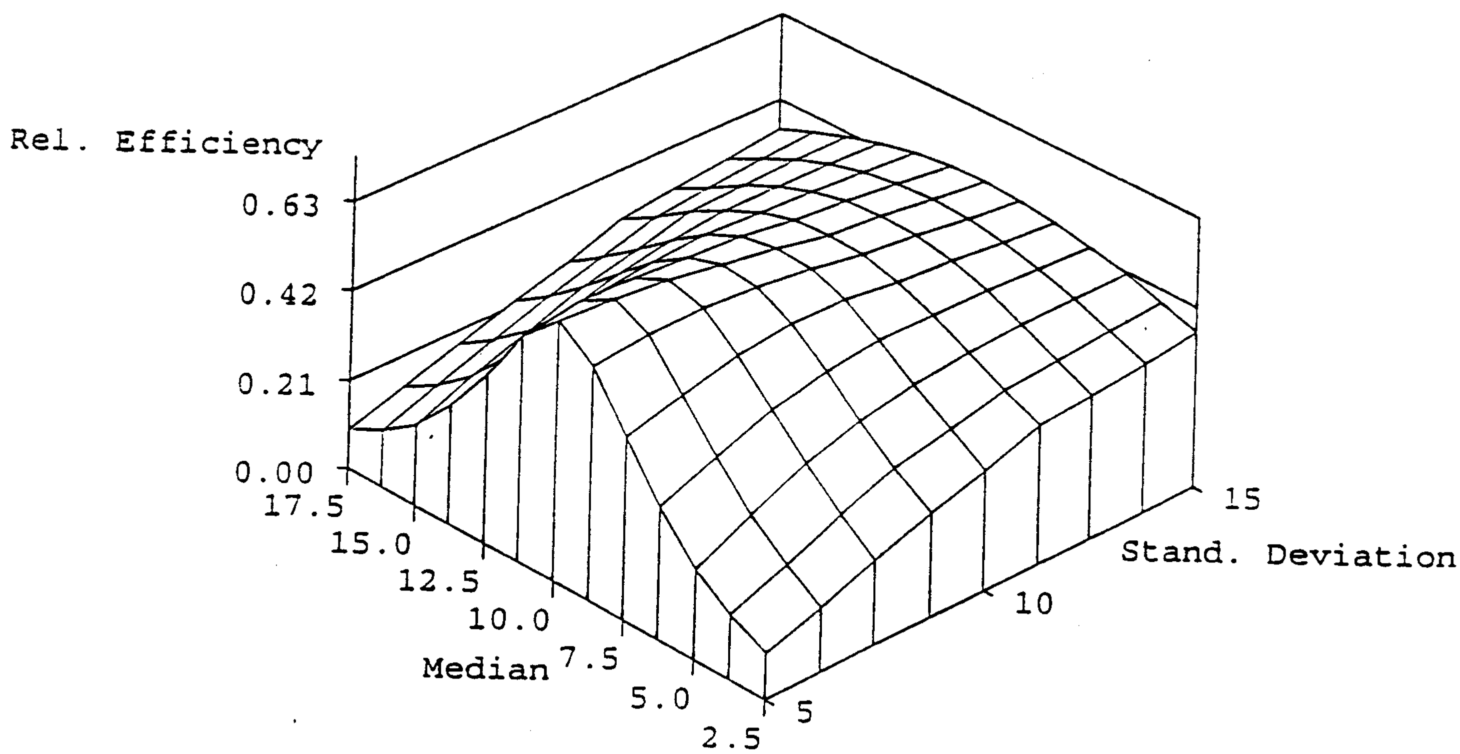
Relative Efficiency of Discrete Data Estimator
Two-point D-optimal Design



Average Relative Efficiencies Over The Replications
As A Function Of The Initial Parameter Estimates
True Value Of Median=10; True Value Of Standard Deviation=10

Figure 5.

Relative Efficiency of Discrete Data Estimator
Quintiles



Average Relative Efficiencies Over The Replications
As A Function Of The Initial Parameter Estimates
True Value Of Median=10; True Value of Standard Deviation=10

When the initial estimate of the median is correct, the fiducial method is quite robust to bad initial variance estimates. The fiducial design's relative efficiency is around 0.60 at the true value of the variance. It improves slightly if the variance is underestimated as its design points move toward the median and drops sharply as the initial estimate of the median departs from its true value. In the latter case, overstating the variance spreads the design points more widely around the assumed median and may push the points far apart enough to place at least one of them relatively close to the true median.

In comparing Figures 3 and 4 note that the relative efficiency of the d-optimal design is much lower than that with the fiducial design when the initial estimates of the parameter values are close to the true ones. For correct guesses concerning μ and σ, the d-optimal approach requires a sample size over 50% larger than the fiducial design for the same relative efficiency. Understating the variance with the d-optimal design tends to result in more efficient designs as the two d-optimal design points are drawn closer to the median. The lowest relative efficiencies for the fiducial method are observed when there is a large bias in the initial estimate for the median and an understated variance. The poor performance of the fiducial method in this situation is related to the placement of its two design points, which are very close to each other and far away from the true median. On the other hand, the lowest relative efficiencies for the d-optimal design occur with a large bias in the initial median estimate and an overstated variance. The poor performance of the d-optimal design in this situation is due to placing one of the two design points far out in one of the tails. The fiducial design dominates the d-optimal design in each of the 7 cells where the median is correctly guessed, and 5 of the cells with a moderate bias and an overstated variance. The d-optimal design dominates in all of the cells where there is a large bias in the initial estimate of the median.

The percentile method in some instances represents a sort of compromise between the fiducial (or the c-optimal) and the d-optimal approaches. While the k=2 percentile design does not have any special merits, being very similar to (if not somewhat inferior to) the fiducial design (see Tables 2.a and 2.c), the percentile designs with a larger number of thresholds, some

TABLE 2a
Average Relative Efficiency of the Estimated Median for Given Initial Estimates for μ and σ.

[in brackets, sample size needed to preserve the efficiency of MLE estimate of the median from 1000 uncensored observations]

Fiducial Method (k even).

BIAS OF THE INITIAL VALUE ASSIGNED TO THE MEDIAN RELATIVE TO THE TRUE VALUE.

Initial value of σ	0%	25%	50%	75%
5.00	0.6289910 [1590]	0.2791580 [3582]	0.0877610 [11395]	0.0392025 [25509]
7.07	0.6126114 [1632]	0.3602740 [2757]	0.1457214 [5862]	0.0679250 [14722]
8.66	0.6057322 [1651]	0.4163159 [2402]	0.1942859 [5147]	0.0951939 [10505]
10.00 [TRUE]	0.6031829 [1658]	0.4412719 [2266]	0.2342360 [4269]	0.1318429 [7585]
11.80	0.5949324 [1681]	0.4697909 [2129]	0.2839385 [3522]	0.1572778 [6358]
12.25	0.5919563 [1689]	0.4754225 [2103]	0.2963593 [3374]	0.1672980 [5977]
13.32	0.5825540 [1717]	0.4891099 [2044]	0.3234109 [3092]	0.1908238 [5240]

TABLE 2b
Average Relative Efficiency of the Estimated Median for Given Initial Estimates for μ and σ.

[in brackets, sample size needed to preserve the efficiency of MLE estimate of the median from 1000 uncensored observations]

D-optimal Design (k even)

BIAS OF THE INITIAL VALUE ASSIGNED TO THE MEDIAN RELATIVE TO THE TRUE VALUE.

Initial value of σ	0%	25%	50%	75%
5.00	0.5661480 [1766]	0.4946159 [2022]	0.3658009 [2734]	0.3012762 [2217]
7.07	0.4816706 [2076]	0.4846736 [2063]	0.4250892 [2352]	0.3437231 [2909]
8.66	0.4360513 [2293]	0.4362835 [2292]	0.4214986 [2372]	0.3826452 [2613]
10.00 [TRUE]	0.3908402 [2559]	0.3896147 [2567]	0.3859160 [2591]	0.3777427 [2647]
11.80	0.3205690 [3119]	0.3207494 [3118]	0.3220435 [3105]	0.3270518 [3057]
12.25	0.3032441 [3298]	0.3030827 [3299]	0.3042571 [3287]	0.3088714 [3237]
13.32	0.2630080 [2802]	0.2621345 [3815]	0.2609878 [3832]	0.2604021 [3840]

TABLE 2c
Average Relative Efficiency of the Estimated Median for Given Initial Estimates for μ and σ.

[in brackets, sample size needed to preserve the efficiency of MLE estimate of the median from 1000 uncensored observations]

33%-percentiles (k=2).

BIAS OF THE INITIAL VALUE ASSIGNED TO THE MEDIAN RELATIVE TO THE TRUE VALUE.

Initial value of σ	0%	25%	50%	75%
5.00	0.6255566 [1766]	0.3096907 [3229]	0.1077331 [9882]	0.0487906 [20495]
7.07	0.6058384 [1651]	0.4015778 [2490]	0.1788827 [5590]	0.0872073 [11467]
8.66	0.6053828 [1652]	0.4414112 [2266]	0.2341617 [4271]	0.1200527 [8330]
10.00 [TRUE]	0.5947947 [1681]	0.4676236 [2139]	0.2788517 [3586]	0.1524575 [6559]
11.80	0.5784141 [1729]	0.4875116 [2051]	0.3301337 [3029]	0.2006705 [4983]
12.25	0.5743022 [1741]	0.4907701 [2037]	0.3425872 [2919]	0.2119675 [4718]
13.32	0.5634338 [1775]	0.4948203 [2021]	0.3687846 [2712]	0.2397943 [4170]

TABLE 2d
Average Relative Efficiency of the Estimated Median for Given Initial Estimates for μ and σ.

[in brackets, sample size needed to preserve the efficiency of MLE estimate of the median from 1000 uncensored observations]

Quintiles (k=4).

BIAS OF THE INITIAL VALUE ASSIGNED TO THE MEDIAN RELATIVE TO THE TRUE VALUE.

Initial value of σ	0%	25%	50%	75%
5.00	0.6023773 [1660]	0.4056219 [2465]	0.1972386 [5070]	0.0988890 [10112]
7.07	0.5882651 [1700]	0.4681341 [2136]	0.2947193 [3393]	0.1684757 [5936]
8.66	0.5580085 [1792]	0.4910441 [2036]	0.3478878 [2875]	0.2263948 [4417]
10.00 [TRUE]	0.5376189 [1860]	0.4951301 [2020]	0.3786436 [2641]	0.2653765 [3768]
11.80	0.5196822 [1924]	0.4863396 [2056]	0.4001910 [2499]	0.2935609 [3406]
12.25	0.5043401 [1983]	0.4791369 [2087]	0.4078903 [2452]	0.3173975 [3151]
13.32	0.4882451 [2048]	0.4677250 [2138]	0.4130149 [2421]	0.3336771 [2997]

TABLE 2e
Average Relative Efficiency of the Estimated Median for Given Initial Estimates for μ and σ.

[in brackets, sample size needed to preserve the efficiency of MLE estimate of the median from 1000 uncensored observations]

14%-percentiles (k=6).

BIAS OF THE INITIAL VALUE ASSIGNED TO THE MEDIAN RELATIVE TO THE TRUE VALUE.

Initial value of σ	0%	25%	50%	75%
5.00	0.6017511 [1662]	0.4167612 [2399]	0.2155374 [4640]	0.1068886 [9356]
7.07	0.5789509 [1727]	0.4764025 [2099]	0.3055655 [3273]	0.1856032 [5388]
8.66	0.5539752 [1805]	0.4920061 [2032]	0.3618820 [2763]	0.2395791 [4174]
10.00 [TRUE]	0.5358414 [1866]	0.4895850 [2042]	0.3913631 [2555]	0.2782665 [3594]
11.80	0.5147538 [1943]	0.4824710 [2073]	0.4027312 [2483]	0.3082509 [3244]
12.25	0.4969901 [2012]	0.4727377 [2115]	0.4106255 [2435]	0.3273440 [3055]
13.32	0.4794147 [2086]	0.4611243 [2169]	0.4115908 [2430]	0.3406947 [2935]

TABLE 2f
Average Relative Efficiency of the Estimated Median for Given Initial Estimates for μ and σ.

[in brackets, sample size needed to preserve the efficiency of MLE estimate of the median from 1000 uncensored observations]

11%-percentiles (k=8).

BIAS OF THE INITIAL VALUE ASSIGNED TO THE MEDIAN RELATIVE TO THE TRUE VALUE.

Initial value of σ	0%	25%	50%	75%
5.00	0.5928510 [1686]	0.4398145 [2511]	0.2348644 [4258]	0.1210007 [8264]
7.07	0.5634282 [1775]	0.4837353 [2067]	0.3319495 [3012]	0.2030062 [4926]
8.66	0.5382317 [1858]	0.4899203 [2041]	0.3758122 [2661]	0.2634850 [3795]
10.00 [TRUE]	0.5148922 [1942]	0.4817124 [2075]	0.3991905 [2502]	0.2977778 [3358]
11.80	0.4945306 [2022]	0.4705580 [2125]	0.4057034 [2464]	0.3230235 [3096]
12.25	0.4747422 [2106]	0.4569732 [2188]	0.4081252 [2450]	0.3379378 [2959]
13.32	0.4569439 [2188]	0.4429341 [2258]	0.4024695 [2485]	0.3446838 [2901]

close to the median and some farther toward the tails, achieve much of the efficiency of the fiducial design when parameter estimates are approximately correct and much of the robustness of the d-optimal design when they are not. For instance, for k equal to 4 the percentile design dominates the fiducial design in 20 out of 28 cells and the d-optimal design in 17 out of 28 cells. The fiducial design continues to offer only a fairly small improvement in relative efficiency if the initial estimate of the median is correct. The d-optimal designs primarily outperform the percentile designs when there is a large bias in the median estimate and the variance is substantially underestimated. Moving from the percentile design with k=2 to the percentile design with k=4 brings about only a small loss of efficiency when the median is correctly guessed, but allows for a marked gain in efficiency with biased initial guesses for the parameters. Further increasing k to 6 improves robustness, albeit at the cost of some efficiency. Increasing k to 8 yields a small gain in robustness while continuing to give up some efficiency at the correct parameter values. Beyond k=8, one seems to gain little, if any at all, additional robustness to bad initial parameter estimates while continuing to give up relative efficiency if the parameter estimates are correct.

The choice between percentile designs with a different number of thresholds depends on the considerations outlined above in regard to robustness (with respect to biased initial estimates) versus efficiency (with good initial estimates). In practice, these considerations will also be weighed against the costs of collecting the data. If increasing the number of distinct design points for a given sample size is virtually costless (as with a computer-assisted telephone interview for which different thresholds can be assigned at almost no extra cost), then one might opt for a relatively large number of thresholds (up to k=8) to achieve a good deal of robustness. If the data are collected by means of personal interviews where the cost of increasing the number of equivalent subsamples (*i.e.*, one for each threshold) for a given sample size is likely to be substantial, then a design with a lower number of thresholds (such as k=4) may be preferred.

5. Concluding remarks

In this paper, we have laid out the basic framework for looking at statistical design issues for binary response data. In doing so, we have drawn heavily on the biometrics literature. We hope that the message conveyed to the reader is that choice of an optimal design for collecting binary data is not a straightforward matter. One must consider specific objectives, available prior information, the likely accuracy of that information, and the ability to sequentially improve that information. The judicious choice of a statistical design can result in dramatic gains in the precision with which model parameters are estimated. We also hope that our presentation will be a useful starting point for econometricians dealing with more complex discrete choice problems. One obvious direction for extending these results is to consider alternative distributional assumptions. Here several frequently used distributions such as the logistic look fairly straightforward. A more difficult direction involves extension to more than two categories of responses. Another direction is to consider the possibility of multiple responses from the same agents obtained in a contemporaneous sequence or over time.

APPENDIX A

Proposition 1. The c-optimal design for μ (if it exists) is a two-point design.

Proof: By Caratheodory's theorem, the c-optimal design for μ, if it exists, requires no more than two design points. We can keep our notation simple by making the assumptions that the two design point are symmetric around the median and that the sample units are equally divided between the two design points.[27] We can write out a formulation of the problem which allows for a one-point design or a two-point design by including a parameter δ which is the distance between the two thresholds, with a distance δ equal to zero meaning a one-point

[27]These two assumptions are not restrictive and are in fact the result of the more general, unrestricted optimization problem. If the c-optimal design is indeed a one-point design, then our requirement that the design points be symmetric around the median means that the optimal design consists of a single threshold that is equal to the median, and that all respondents will be assigned that value of the threshold.

design that concentrates the thresholds at the median μ.[28] For the optimal design to consist of one threshold, it is sufficient to show that the MSE attains a minimum at δ=0. For the c-optimal design to consist of two points, it is necessary to prove that the MSE does *not* attain a minimum at δ=0, hence the optimal δ (if it exists) must be > 0.

The $MSE_{[r]}$, the asymptotic mean square error based on the r-th order expansion approximation to the distribution of $\sqrt{n}(\hat{\mu}-\mu)$, is written:

$$MSE_{[r]} = 4\delta^2\left\{\frac{2\theta^2}{(2\beta\delta)^4}v(\theta)+\frac{2\theta^2\Gamma(r+1/2)}{\sqrt{\pi}}\frac{v(\theta)^r}{n^{r-1}(2\beta\delta)^{2(r+1)}}+6\sum_{k=2}^{r-1}\frac{2^k\theta^2\Gamma(k+1/2)}{\sqrt{\pi}}\frac{v(\theta)^k}{n^{k-1}(2\beta\delta)^{2(k+1)}}\right\}$$

for $r\geq 2$, and $MSE_{[1]} = 4\delta^2\left\{\frac{2(\beta\delta)^2}{(2\beta\delta)^4}v(\theta)\right\}$ for $r=1$, where $\theta=\beta\delta$ and $v(\theta)=2\Phi(\theta)[1-\Phi(\theta)]/\phi^2(\theta)$.

We will show that $\lim_{\delta\to 0}\left[MSE_{[r]}\right]_{r\to\infty} > \inf_{\delta}\left[MSE_{[r]}\right]_{r\to\infty}$, hence the c-optimal design, if it exists, cannot be a one-point design, and must be a two-point design.

Because each of the terms entering in the expression for the MSE is easily verified as non-negative, $\lim_{\delta\to 0}\left[MSE_{[r]}\right]_{r\to\infty} \geq \lim_{\delta\to 0}\left[MSE_{[r]}\right]_{\text{any finite } r}$. In order to prove that the c-optimal design must be a two-point design, we must show that $\lim_{\delta\to 0}\left[MSE_{[r]}\right]_{\text{any finite } r} > \inf_{\delta}\left[MSE_{[r]}\right]_{r\to\infty}$.

It will be easiest to prove the latter inequality for r=2. The expression for the $MSE_{[2]}$ is:

$$MSE_{[2]} = 4\delta^2\left\{\frac{2\theta^2}{(2\beta\delta)^4}v(\theta)+\frac{2\theta^2\Gamma(2+1/2)}{\sqrt{\pi}}\frac{v(\theta)^2}{n(2\beta\delta)^6}\right\}.$$

On taking the limit for fixed sample size n,

$$\lim_{\delta\to 0} MSE_{[2]} = \lim_{\delta\to 0}\left\{4\delta^2\left[\frac{2\theta^2}{(2\beta\delta)^4}v(\theta)+\frac{2\theta^2\Gamma(2+1/2)}{\sqrt{\pi}}\frac{v(\theta)^2}{n(2\beta\delta)^6}\right]\right\} =$$

---

[28]Ford, Torsney and Wu (1992) have argued that the c-optimal design for μ is a one-point design.

$$= \lim_{\delta\to 0}\left\{4\delta^2\left[\frac{2(\beta\delta)^2}{(2\beta\delta)^4}\nu(\theta)+\frac{2(\beta\delta)^2\Gamma(2+1/2)}{\sqrt{\pi}}\frac{\nu(\theta)^2}{n(2\beta\delta)^6}\right]\right\}=\frac{1}{2}\frac{\nu(0)}{\beta^2}\lim_{\delta\to 0}\frac{\delta^4}{\delta^4}+\frac{1}{4}\frac{\nu(0)^2\Gamma(5/2)}{\sqrt{\pi}\beta^4}\lim_{\delta\to 0}\frac{\delta^4}{n\delta^6}.$$

This is is a 0/0 indeterminate form. The limit can be calculated by rewriting this quantity as

$$\lim_{\delta\to 0} MSE_{[2]}=\frac{1}{2}\frac{\nu(0)}{\beta^2}\lim_{\xi\to\infty}\frac{(1/\xi)^4}{(1/\xi)^4}+\frac{1}{4}\frac{\nu(0)^2\Gamma(5/2)}{\sqrt{\pi}\beta^4}\lim_{\xi\to\infty}\frac{(1/\xi)^4}{n(1/\xi)^6}$$

and applying De L'Hospital rule repeatedly to obtain:

$$\lim_{\delta\to 0} MSE_{[2]}=\frac{1}{2}\frac{\nu(0)}{\beta^2}+\frac{1}{4}\frac{\nu(0)^2\Gamma(5/2)}{\sqrt{\pi}\beta^4}\lim_{\xi\to\infty}\frac{(1/\xi)^4}{n(1/\xi)^6}=\frac{1}{2}\frac{\nu(0)}{\beta^2}+\frac{1}{4}\frac{\nu(0)^2\Gamma(5/2)}{\sqrt{\pi}\beta^4}(+\infty)=+\infty.$$

Therefore, while the first-order term converges to a finite limit, $\frac{1}{2\beta^2}\nu(0)\cong\frac{2\cdot 1.57}{2\beta^2}=1.57\sigma^2$ as the two design points get closer and closer to the center, the second-order terms explodes.[29] Hence the MSE, either for r=2 or for a larger order of the expansion[30], cannot attain a finite

---

[29] $1.57\sigma^2$ is in fact the lower bound for $MSE_{[1]}$.

[30] It can be shown that the terms of order higher than two grow towards infinity at a faster rate than the second-order term as $\delta\to 0$. The third-order term, for instance, is:

$$\kappa_3=4\delta^2\left\{\frac{2\theta^2\Gamma(3+1/2)}{\sqrt{\pi}}\frac{\nu(\theta)^3}{n^2(2\beta\delta)^8}\right\}.$$

If we take the ratio of this term over the second-order term,

$$\kappa_2=4\delta^2\left\{\frac{2\theta^2\Gamma(2+1/2)}{\sqrt{\pi}}\frac{\nu(\theta)^2}{n\ (2\beta\delta)^6}\right\},$$

we obtain $\kappa_3/\kappa_2=\dfrac{4\delta^2\left\{\frac{2\theta^2\Gamma(3+1/2)}{\sqrt{\pi}}\frac{\nu(\theta)^3}{n^2(2\beta\delta)^8}\right\}}{4\delta^2\left\{\frac{2\theta^2\Gamma(2+1/2)}{\sqrt{\pi}}\frac{\nu(\theta)^2}{n\ (2\beta\delta)^6}\right\}}$. For large δ this ratio is simplified to $\frac{\Gamma(3+1/2)}{\Gamma(2+1/2)}\frac{\nu(\theta)}{n(2\beta\delta)^2}$ and tends to zero even for relatively small sample sizes (the higher-order terms vanish for sufficiently large sample size and sufficiently large finite δ). If $\delta\to 0$ the limit of this quantity is infinity.

value as $\delta \to 0$. This proves that the c-optimal design point, if it exists, is a two-point design ($\delta > 0$). It also proves that $\lim_{\delta \to 0} MSE_{[1]} \neq \left[\lim_{\delta \to 0} MSE_{[r]}\right]_{r \to \infty}$: the MSE based on the first-order expansion approximation and the MSE based on a higher-order expansion approximation tend to different limits as $\delta \to 0$, and thus that the distribution of $\sqrt{n}(\hat{\mu} - \mu)$ is not adequately approximated using a first-order Taylor series approximation. A one-point design thus violates the requirement for adequacy of the first-order approximation that the higher-order terms should vanish.

APPENDIX B.

Fiducial design for each conditional median in a model with regressors.

Consider a model with covariates (Section 3.5) and suppose we choose to minimize the squared half length of the fiducial interval around each median.[31] Assume that the individual values of the regressors are known prior to the study, so that the objective function for each m_j can be formulated conditionally on the knowledge of such values. Adapting the expression for the squared half-length of the fiducial interval around the median (Section 3.2) to the model with covariates, we obtain:

$$(9)\quad \frac{t^2}{\hat{\beta}_c}\left[Var(x_j\hat{\beta}) + 2\left[-\frac{x_j\hat{\beta}}{\hat{\beta}_c}\right]Cov(x_j\hat{\beta},\hat{\beta}_c) + \left[-\frac{x_j\hat{\beta}}{\hat{\beta}_c}\right]^2 Var(\hat{\beta}_c) - g\left[Var(x_j\hat{\beta}) - \frac{Cov^2(x_j\hat{\beta},\hat{\beta}_c)}{Var(\hat{\beta}_c)}\right]\right](1-g)^{-2},$$

where $g = \dfrac{t^2 Var\left(\hat{\beta}_c\right)}{\hat{\beta}_c^2}$, t is the value of the standard normal deviate corresponding to the desired fiducial level, $\beta = \gamma / \sigma_\varepsilon$ and $\beta_c = -1/\sigma_\varepsilon$. $Var\left(x_j\hat{\beta}\right)$ and $Cov(x_j\hat{\beta},\hat{\beta}_c)$ can be written

---

Hence, the terms of order higher than two grow towards infinity at rate faster than the second-order term does as δ tends to zero, but vanish faster than the second-order term does for sufficiently large but finite δ.

[31]The thresholds for the d-optimal design for the case including covariates can be derived in a similar manner, as the criterion function there, like with the fiducial method, is defined on the information matrix for the parameters.

as $x_j Var(\hat{\beta})x_j^{'}$ and $x_j Cov(\hat{\beta},\hat{\beta}_c)$ respectively. The expression for the covariance matrix of $(\hat{\beta},\hat{\beta}_c)'$ is:

$$I(\beta,\beta_c)=\left[\sum_{i=1}^{n}\frac{\phi^2(z_i)}{\Phi(z_i)[1-\Phi(z_i)]}\begin{bmatrix} x_i x_i^{'} & x_i c_i \\ c_i x_i^{'} & c_i^2 \end{bmatrix}\right]^{-1},$$

where $z_i = x_i\beta+\beta_c c_i$. Equation (9) shows the optimal thresholds depend on the individual realizations of the covariates, plus γ and σ_ε. The objective function (9) can be rewritten in the following fashion:

$$(10)\quad t^2\sigma^2\left[Var\left(\frac{\hat{m}_j}{\hat{\sigma}}\right)+2\hat{m}Cov\left(\frac{\hat{m}_j}{\hat{\sigma}},\frac{1}{\hat{\sigma}}\right)+\hat{m}_j^2 Var\left(\frac{1}{\hat{\sigma}}\right)-g\left[Var\left(\frac{\hat{m}_j}{\hat{\sigma}}\right)-\frac{Cov^2\left(\frac{\hat{m}_j}{\hat{\sigma}},\frac{1}{\hat{\sigma}}\right)}{Var(1/\hat{\sigma})}\right]\right](1-g)^{-2}$$

where $\hat{m}_j$ indicates the conditional median of $y_j^*, x_i\gamma$. From equation (10), one can compute the optimal thresholds using complete information on the conditional first (m_i) and the second (σ^2) moments of the y_i^* which incorporate all the information about the individual x_i^*'s.

References

Abdelbasit, K.M. and R.L. Plackett, 1983, Experimental design for binary data, Journal of the American Statistical Association, 78, 90-98.

Aigner, D. and C. Morris, eds., 1979, Experimental design in econometrics, special issue of the Journal of Econometrics, 11.

Alberini, A. and R.T. Carson, 1991, The informational content of discrete choice responses, draft manuscript, Department of Economics, University of California, San Diego, June.

Amemiya, T., 1985, Advanced Econometrics (Harvard University Press, Cambridge).

Arrow, K.J. L. Pesotchinsky and M. Sobel, 1981, On partitioning a sample with binary-type questions in lieu of collecting observations, Journal of the American Statistical Association, 76, 402-409.

Cameron, T.A., 1988, A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression, Journal of Environmental Economics and Management, 15, 355-379.

Cameron, T.A. and M.D. James, 1987a, Efficient estimation methods for use with 'closed ended' contingent valuation survey data, Review of Economics and Statistics, 69, 269-276.

Cameron, T.A. and M.D. James, 1987b, Estimating willingness to pay form survey data: An alternative pre-test market evaluation procedure, Journal of Marketing Research, 24, 389-395.

Chernoff, H (1979), Sequential Analysis and Optimal Design (SIAM, Philadelphia).

Conlisk, J., 1973, Choice of response function in designing subsidy experiments, Econometrica, 41, 643-656.

Duffield, J.W. and D.A. Patterson (1991), Inference and optimal design for a welfare measure in dichotomous choice contingent valuation, Land Economics, 67, 225-239.

Ferber, R. and W.Z. Hirsch, 1982, Social experimentation and economic policy (Cambridge University Press, New York).

Finney, D.J., 1978, Statistical methods in biological assays, 3rd ed. (Oxford University Press, New York).

Fisher, R.A., 1956, Statistical methods and scientific inference (Oliver & Boyd, Edinburgh).

Ford, I. and S.D. Silvey, 1980, A sequentially constructed design for estimating a non-linear parametric function, Biometrika, 67, 381-388.

Ford, I., B. Torsney, and C.F.J. Wu, 1992, The use of canonical form in the construction of locally optimal designs for nonlinear problems, Journal of the Royal Statistical Society, Series B, 54, 569-583.

Frey, J.H., 1989, Survey research by telephone, 2nd ed. (Sage Publications, Newbury Park).

Kanninen, B.J., 1991, Optimal Experimental Designs for Contingent Valuation Surveys, Ph.D. dissertation, University of California, Berkeley.

Kanninen, B.J., 1993, Design of Sequential Experiments for Contingent Valuation Studies, Journal of Environmental Economics and Management, 25, S-1-S-11.

Lehmann, E.L., 1983, Theory of point estimation (New York: John Wiley).

Manski, C. and S. Lerman, 1977, The estimation of choice probabilities from choice-based samples, Econometrica, 45, 1977-1988.

McFadden, D., 1981, Econometric models of probabilistic choice, in: C.F. Manski, and D. McFadden, eds., Structural analysis of discrete data with econometric applications (Cambridge, MIT Press).

McLeish, D.L. and D. Tosh, 1990, Sequential designs in bioassays, Biometrics, 46, 103-116.

Minkin, S., 1987, Optimal design for binary data, Journal of the American Statistical Association, 82, 1098-1103.

Mitchell, R.C. and R.T. Carson, 1989, Using surveys to value public goods: the contingent valuation method (Resources for the Future/Johns Hopkins University Press, Washington).

Nyquist, H., 1992, Optimal designs of discrete response experiments in contingent valuation studies, The Review of Economics and Statistics, 74, 559-563.

Silvey, S.D., 1980, Optimal design (Chapman & Hall, New York).

Smith, V.L., 1980, Experiments with a decentralized mechanism for public good decisions, American Economic Review, 70, 584-599.

Tsutakawa, R.K., 1972, Design of experiments for bioassay, Journal of the American Statistical Association, 67, 584-590.

Tsutakawa, R.K., 1980, Selection of dose levels for estimating a percentage point of a logistic quantal response curve, Applied Statistics, 29, 25-33.

Wetherhill, G.B., 1975, Sequential methods in statistics (Chapman and Hall, London).

White, H., 1984, Asymptotic Theory for Econometricians, (Academic Press, Orlando).

Wu, C.F.J., 1988, Optimal design for percentile estimation of a quantal response curve, in Y. Dodge, V.V. Fedorov, and H.P. Wynn, eds., Optimal design and analysis of experiments (Elsevier Science Publishers, Amsterdam).