RESEARCH REPORT

Serial No. 31

December 1992

BAYESIAN OPTIMAL DESIGNS FOR PROBIT REGRESSION WITH ERRORS-IN-VARIABLES

by

P.K. Tang and J. Bacon-Shone

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS

Pokfulam Road, Hong Kong. Tel: (852) 859 2466 Fax: (852) 858 9041
E-Mail: STATIST@HKUCC.HKU.HK or STATIST@HKUCC.BITNET

BAYESIAN OPTIMAL DESIGNS FOR PROBIT REGRESSION WITH ERRORS-IN-VARIABLES

P K Tang

Securities Relationship Division, HongkongBank, L17, 1 Queen's Road Central, Hong Kong.

John Bacon-Shone

Social Sciences Research Centre and Department of Statistics, University of Hong Kong, Pokfulam Road, Hong Kong.

Summary

Optimal design is the study of the choice of design points in an experiment. However, measurements are seldom precise in practical situations. If measurement error is substantial, it may ruin the whole experiment in that the objective of the experiment is not achieved. There is substantial literature on optimal designs, all based on the assumption that there is no measurement error in the covariates. For the Berkson error model, the observed design points are fixed by the experimenter but they deviate randomly from the pre-assigned level. In this paper, the Berkson error structure is incorporated into the probit regression model for which Bayesian D-optimal and A-optimal designs are studied. In addition, a new optimal design criterion is proposed.

Keywords: optimal design, posterior mode, errors-in-variables, Berkson's model, probit regression.

1 Introduction

Suppose that the experiment is to be designed by choosing n values of the design variable x from an experimental region X. Let the unknown parameters be $\theta^T = (\theta_1,\ldots,\theta_p)$. By expanding the definition of a design to include any probability measure η on Z we define the normalized matrix $I(\hat{\theta},\eta)$ by

$$[I(\hat{\theta},\eta)]_{ij} = -\int E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log(p(y \mid \theta, x))\right]_{\theta=\hat{\theta}} \eta(dx).$$

Whittle (1973) stated a necessary condition, which involves a directional derivative, to check if the design for linear regression model without measurement error is optimal. An explanation of directional derivative is given as follows:

For two measures η_1 and η_2 in H the derivative at η_1 in the direction of η_2 is defined by

$$F(\eta_1,\eta_2) = \lim_{\epsilon \to 0} \frac{\phi((1-\epsilon)\eta_1 + \epsilon\eta_2) - \phi(\eta_1)}{\epsilon} .$$

The special case that η_2 is a one point design plays a key role in Whittle's findings. We will denote such $F(\eta_1,\eta_2)$ by $d(\eta,x)$. Whittle (1973) proved that if η_0 is the optimal design then the support points will be the roots of $d(\eta_0, x) = 0$.

Chaloner and Larntz (1989) gave the conditions under which Whittle's condition applies to non-linear regression models. The assumptions are:

1. X is compact;
2. the directional derivative exists and is continuous in X;
3. there is at least one measure in H, where H is the set of all probability measures on X, for which ϕ is finite; and
4. ϕ is such that if $\eta_i \to \eta$ in weak convergence then $\phi(\eta_i) \to \phi(\eta)$.

Note that X in the assumptions refers to the experimental region. Under the Berkson error model, the proxy variable, z, which can be controlled by the experimenter, is observed rather than the design variable x. Berkson's measurement error model is particularly suitable in this situation. Therefore, the experiment turns into one designed by choosing n values of the observed variable z from a controllable region, Z. In practice, we are dealing with the controllable region and so assume the above assumptions also hold for Z. Hence, Whittle's condition can be applied in our study.

Given the prior distribution of θ, we would like to find the designs that are derived from the maximization of two concave criteria which were studied by Chaloner and Larntz (1989). The first criterion, which corresponds to maximizing the average over the prior distribution of the log of the determinant of the expected information matrix, is

$$\phi_1(\eta) = E_\theta \log | I(\theta, \eta) |.$$

For a design measure η for which $I(\theta,\eta)$ is singular for θ values of non-zero prior probability we define ϕ_1 to be $-\infty$. This criterion corresponds to the D-optimality criterion in the linear case.

The second criterion corresponds to minimizing the approximate expected posterior variance of the quantities of interest. This criterion requires the experimenter to specify what is to be estimated or what is to be predicted and the relative importance of these predictions. A weighted trace of the inverse of the expected information matrix is then averaged over the prior distribution and minimized. This criterion corresponds to A-optimality in linear design and it is stated as the maximization of

$$\phi_2(\eta) = -E (tr (B(\theta) I(\theta, \eta)^{-1})).$$

The relative importance of the predictions is specified through the symmetric matrix B.

The corresponding directional derivatives for the two criteria are

$$d(\eta, z) = E_\theta \, tr \, (I(\theta,z) \, I(\theta,\eta)^{-1}) - p \; ;$$

and

$$d(\eta,z) = E_\theta \, tr \, (B(\theta) \, I(\theta,\eta)^{-1} \, I(\theta,z) \, I(\theta,\eta)^{-1}) + \phi_2(\eta) \; .$$

For most models, the posterior distribution is intractable and asymptotic arguments are used. The asymptotic posterior distribution of the Bayes estimator is specified in terms of either the observed Fisher's information matrix or the expected Fisher's information matrix. For generalized linear models with canonical link function, the two matrices are the same. For generalized linear models with non-canonical link function, the observed Fisher's information matrix is different from the expected Fisher's information matrix. For simplicity, the expected Fisher's information matrix is used, although Efron and Hinkley (1978) have shown that, in the case of MLE, if the likelihood function is close to normal then the variance approximation by the observed Fisher's information is better than by the expected Fisher's information.

Assuming that the expected information matrix is non-singular, the posterior distribution of θ using a design measure η is approximately distributed as

$$N_p \left(\hat{\theta}, \, nI(\hat{\theta}, \, \eta)^{-1} \right)$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ.

2 Probit Model

Recall that the p.d.f. of y=1 conditional on z (Burr, 1988) is given by

$$\begin{aligned} p(y=1 \mid z) &= \int \Phi(\beta(z-\gamma)\phi\left(\frac{x-z}{\sigma_u}\right) dx \\ &= \Phi\left(\frac{\beta}{\sqrt{1+\beta^2\sigma_u^2}} (z-\gamma) \right) \\ &= \Phi(\beta_r(z-\gamma)\,) \end{aligned}$$

where

$$\beta_r = \frac{\beta}{\sqrt{1 + \beta^2 \sigma_u^2}}$$

since, under the Berkson error model,

$$X = Z + U, \quad U \sim N(0, \sigma_u^2).$$

It is further assumed that σ_u^2 is known.

Here, $\theta = (\gamma, \beta)$. The likelihood for γ and β is then

$$L = \prod_{i=1}^{n} \left[\Phi\left(\frac{\beta(z_i - \gamma)}{\sqrt{1+\beta^2\sigma_u^2}} \right) \right]^{y_i} \left[1 - \Phi\left(\frac{\beta(z_i - \gamma)}{\sqrt{1+\beta^2\sigma_u^2}} \right) \right]^{1-y_i}.$$

The expected Fisher's information matrix for γ and β is

$$I = \begin{bmatrix} \sum_{i=1}^{n} \frac{1}{\Phi_i(1-\Phi_i)} \left(\frac{\partial \Phi_i}{\partial \gamma} \right)^2 & \sum_{i=1}^{n} \frac{1}{\Phi_i(1-\Phi_i)} \left(\frac{\partial \Phi_i}{\partial \gamma} \right) \left(\frac{\partial \Phi_i}{\partial \beta} \right) \\ \sum_{i=1}^{n} \frac{1}{\Phi_i(1-\Phi_i)} \left(\frac{\partial \Phi_i}{\partial \gamma} \right) \left(\frac{\partial \Phi_i}{\partial \beta} \right) & \sum_{i=1}^{n} \frac{1}{\Phi_i(1-\Phi_i)} \left(\frac{\partial \Phi_i}{\partial \beta} \right)^2 \end{bmatrix}$$

where

$$\frac{\partial \Phi_i}{\partial \gamma} = -\frac{\beta}{\sqrt{1+\beta^2\sigma_u^2}} \, \phi\left(\frac{\beta(z_i - \gamma)}{\sqrt{1+\beta^2\sigma_u^2}} \right);$$

and

$$\frac{\partial \Phi_i}{\partial \beta} = \frac{z_i - \gamma}{(1+\beta^2\sigma_u^2)^{\frac{3}{2}}} \, \phi\left(\frac{\beta(z_i - \gamma)}{\sqrt{1+\beta^2\sigma_u^2}} \right).$$

3 Classical D-optimal Design

In the error free situation, the classical approach to a two point D-optimal design would mean that given γ and β, the support points are symmetrically placed on either side of γ (Abdelbasit and Plackett, 1981 and Minkin, 1987). Let c = z - γ and the design criterion is to maximize

$$| I | = \frac{n^2 \beta^2 c^2}{\Phi^2 (1 - \Phi)^2} \phi^4(\beta c), \tag{1}$$

where

$$\Phi = \Phi(\beta c).$$

In the current situation, the classical approach would find c that maximizes (1) with β replaced by β_r.

Let $p^* = \Phi(\beta_r c)$ where c maximizes (1). Then

$$\Phi^{-1}(p^*) = \frac{\beta}{\sqrt{1 + \beta^2 \sigma_u^2}} (z - \gamma)$$

i.e.

$$z = \frac{\Phi^{-1}(p^*) \sqrt{1 + \beta^2 \sigma_u^2}}{\beta} + \gamma. \tag{2}$$

It can be seen that z is an increasing function in σ_u and a decreasing function in β. Hence, we need to put the support points further away from γ if error is present.

Given a known value of β, β_0, c is determined from (1). Then

$$p = \Phi(\beta_r (z-\gamma))$$

$$= \Phi\left(\frac{\beta}{\sqrt{1+\beta^2\sigma_u^2}}\left(\frac{\Phi^{-1}(p^*)\sqrt{1+\beta_0^2\sigma_u^2}}{\beta_0}\right)\right) \qquad (3)$$

Consider the behaviour of p when $\beta \to \infty$

$$\lim_{\beta\to\infty} p = \lim_{\beta\to\infty} \Phi\left(\frac{\beta}{\sqrt{1+\beta^2\sigma_u^2}}\left(\frac{\Phi^{-1}(p^*)\sqrt{1+\beta_0^2\sigma_u^2}}{\beta_0}\right)\right)$$

$$= \Phi\left(\frac{\Phi^{-1}(p^*)\sqrt{1+\beta_0^2\sigma_u^2}}{\sigma_u\beta_0}\right)$$

It can be seen that the conditional density does not go to 0 or 1 as β goes to infinity which causes an estimation problem (Tang and Bacon-Shone, 1992). Furthermore, if error is ignored at the design stage, i.e. set $\sigma_u=0$ in (2), the likelihood does not decay to zero and it is thicker in the tail than when error is taken into consideration at the design stage. In fact, the design resulting from ignoring error at the design stage will result in a likelihood with the thickest tail and hence there is the greatest chance of getting an infinite maximum likelihood estimate of β.

4 Bayesian Optimal Designs

We consider Bayesian optimal designs under 3 criteria. The first criterion, ϕ_1, is the D-optimal criterion, which is same as the Fisher's information criterion. The second criterion, ϕ_2, is the A-optimal criterion which is derived from using the posterior mean as the Bayes point estimate. The third criterion, ϕ_3, is the optimal criterion derived from using the posterior mode as the Bayes point estimate as it has been recommended that the posterior mode should be used for estimation in binary regression models with measurement error (Tang and Bacon-Shone, 1992).

Given a design measure, η, putting n_i weight at k distinct design points z_i, $i=1,\ldots,k$, $\sum n_i = 1$, we define

$$\phi_i = \phi\left(\frac{\beta}{\sqrt{1+\beta^2\sigma_u^2}}(z_i-\gamma)\right); \quad \Phi_i = \Phi\left(\frac{\beta}{\sqrt{1+\beta^2\sigma_u^2}}(z_i-\gamma)\right)$$

$$w_i = \frac{\phi_i^2}{\Phi_i(1-\Phi_i)}; \qquad t = \sum_{i=1}^{k} n_i\, w_i$$

$$\bar{z} = \frac{\sum_{i=1}^{k} n_i\, w_i\, z_i}{\sum_{i=1}^{k} n_i\, w_i}; \qquad s = \sum_{i=1}^{k} n_i\, w_i\,(z_i-\bar{z})^2 .$$

Then, the expected Fisher's information matrix can be re-expressed as

$$I(\theta,\eta) = \begin{bmatrix} \dfrac{\beta^2 t}{1+\beta^2\sigma_u^2} & -\dfrac{\beta t(\bar{z}-\gamma)}{(1+\beta^2\sigma_u^2)^2} \\ -\dfrac{\beta t(\bar{z}-\gamma)}{(1+\beta^2\sigma_u^2)^2} & \dfrac{s+t(\bar{z}-\gamma)^2}{(1+\beta^2\sigma_u^2)^3} \end{bmatrix}$$

And the determinant of the expected Fisher's information matrix becomes

$$|I(\theta,\eta)| = \frac{\beta^2 ts}{(1+\beta^2\sigma_u^2)^4}$$

The inverse of $I(\eta)$ is

$$I^{-1}(\theta,\eta) = \begin{bmatrix} \left(\dfrac{1}{t}+\dfrac{(\bar{z}-\gamma)^2}{s}\right)\dfrac{(1+\beta^2\sigma_u^2)}{\beta^2} & \dfrac{(\bar{z}-\gamma)(1+\beta^2\sigma_u^2)^2}{s\beta} \\ \dfrac{(\bar{z}-\gamma)(1+\beta^2\sigma_u^2)^2}{s\beta} & \dfrac{(1+\beta^2\sigma_u^2)^3}{s} \end{bmatrix}.$$

4.1 ϕ_1 Optimal Design

It can be shown that the criterion function and the directional derivative for the ϕ_1 optimal design are respectively

$$\phi_1(\eta) = E_\theta\left[\log\left(\frac{\beta^2 ts}{(1+\beta^2\sigma_u^2)^4}\right)\right];$$

and

$$d(\eta,z) = E_\theta\left[w\left(\frac{1}{t}+\frac{(\bar{z}-z)^2}{s}\right)\right] - 2 .$$

Note that the directional derivative function is in the same form as that of the no error case. As $|z| \to \infty$, $w \to 0$ at an exponential rate, i.e. $d(\eta,z) \to -2$. Hence, the support points will never be placed at infinity. This is also in accordance with the no error situation.

4.2 ϕ_2 Optimal Design

In the case of ϕ_2 optimality, the experimenter is required to specify what is to be estimated or predicted. Suppose we wish to estimate z_0 such that $\Phi(\beta_r(z_0-\gamma)) = \lambda$. We have

$$z_0 = \frac{\Phi^{-1}(\lambda)}{\beta_r} + \gamma.$$

This expression is non-linear in the parameters. The asymptotic variance of $\hat{z}_0$ can be found by using the delta method (Serfling, 1980, p. 118). Thus, $B(\theta)$ can be defined as

$$B(\theta) = c(\theta)c(\theta)^T$$

where

$$c(\theta)^T = \left(1,\ \frac{-\Phi^{-1}(\lambda)}{\beta^2\sqrt{1+\beta^2\sigma_u^2}}\right).$$

Therefore, we have the criterion function and directional derivative of ϕ_2 optimality for the percentiles response point to be

$$\phi_2(\eta) = -E_\theta\left\{\frac{1+\beta^2\sigma_u^2}{\beta^2}\left[\frac{1}{t}+\frac{1}{s}\left(\bar{z}-\gamma-\frac{\Phi^{-1}(\lambda)\sqrt{1+\beta^2\sigma_u^2}}{\beta}\right)^2\right]\right\};$$

and

$$d(\eta,z)$$

$$= E_\theta\left\{\frac{w}{\beta^2}\left[\sqrt{1+\beta^2\sigma_u^2}\left(\frac{1}{t}+\frac{(\bar{z}-\gamma)(\bar{z}-z)}{s}\right)-\frac{\Phi^{-1}(\lambda)(\bar{z}-z)(1+\beta^2\sigma_u^2)}{s\beta}\right]^2\right\}$$

$$+ \phi_2(\eta)$$

respectively.

Note that if we want to estimate γ alone, we can set $\lambda = 0.5$. If both γ and β are of equal interest, $B(\theta)$ is the identity matrix. For this $B(\theta)$, we have the criterion function and the directional derivative function of ϕ_2 optimal design for the sum of parameters to be

$$\phi_2(\eta) = -E_\theta\left\{\left(\frac{1}{t}+\frac{(\bar{z}-\gamma)^2}{s}\right)\frac{(1+\beta^2\sigma_u^2)}{\beta^2} + \frac{(1+\beta^2\sigma_u^2)^3}{s}\right\}$$

and

$$d(\eta,z) = \frac{w(1+\beta^2\sigma_u^2)}{\beta^2}\left(\frac{1}{t}+\frac{(\bar{z}-\gamma)(\bar{z}-z)}{s}\right)^2 + \frac{w(\bar{z}-z)^2(1+\beta^2\sigma_u^2)^3}{s^2}$$

Note that there is singularity at $\beta=0$ in the objective functions.

4.3 ϕ_3 Optimal Design

If $\tilde{\theta}$ is the posterior mode, then, as the sample size gets large,

$$\tilde{\theta} \sim N\left(\hat{\theta}\ ,\ [I(\hat{\theta})]^{-1}\right)$$

where $\hat{\theta}$ is the MLE and $I(\theta)$ is the expected Fisher information matrix with (i,j) element

$$I_{i,j} = -nE\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(z\,|\,\theta)\right]$$

To minimize the loss w.r.t. the 0-1 loss function is essentially the same as to maximize the p.d.f. of the posterior mode. From the asymptotic theory, the mode of the posterior distribution is the MLE and the corresponding density is inversely proportional to the square root of the determinant of the variance-covariance matrix. The variance-covariance matrix is the inverse of the Fisher's information matrix evaluated at the MLE. Therefore, the design criterion of the posterior mode, call it the ϕ_3 criterion, becomes the maximization of the square root of the determinant of the expected

Fisher's information matrix. It seems that there is no reported research on optimal design for the posterior mode.

For ϕ_3-optimality, i.e. the average of the square root of the determinant of the Fisher's information matrix is maximized, the criterion function and the directional derivative function are respectively

$$\phi_3(\eta) = E_\theta\left[\frac{\beta^2 ts}{(1+\beta^2\sigma_u^2)^4}\right]^{\frac{1}{2}}.$$

and

$$d(\eta,z) = E_\theta\left\{\frac{\beta t^{\frac{1}{2}} s^{\frac{1}{2}}}{2(1+\beta^2\sigma_u^2)^2}\left[w\left(\frac{1}{t}+\frac{(\bar{z}-z)^2}{s}\right)-2\right]\right\}$$

where w is the design point of a single point design. Note that as $|z| \to \infty$, $w \to 0$ at the exponential rate and hence $d(\eta,z) \to$ constant. Therefore, the support points will never be placed at infinity. The derivation of the criterion and derivative functions are given in the appendix.

4.4 Tail behaviour of the design criteria

Let us look at the tail behaviour of these design criteria. For a given z_i, when $\beta \to \infty$,

$$\phi_i \to \phi\left(\frac{z_i-\gamma}{\sigma_u}\right);$$

$$\Phi_i \to \Phi\left(\frac{z_i-\gamma}{\sigma_u}\right);$$

$$w_i \to \frac{\phi_i^2\left(\frac{z_i-\gamma}{\sigma_u}\right)}{\Phi_i\left(\frac{z_i-\gamma}{\sigma_u}\right)\left(1-\Phi_i\left(\frac{z_i-\gamma}{\sigma_u}\right)\right)};$$

We can see that w_i tends to a constant as $\beta \to \infty$. This behaviour of the w_i's will make t, s and $\bar{z}$ tend to constants as β goes to infinity.

The criterion function of the ϕ_1 optimal design is

$$\phi_1(\eta) = E_\theta\left[\log\left(\frac{\beta^2 ts}{(1+\beta^2\sigma_u^2)^4}\right)\right].$$

The argument of the log function goes to 0 as β goes to infinity.

Recall that the design criterion of ϕ_2 optimal design for the percentile response points is given by

$$\phi_2(\eta) = -E_\theta\left\{\frac{1+\beta^2\sigma_u^2}{\beta^2}\left[\frac{1}{t}+\frac{1}{s}\left(\bar{z}-\gamma-\frac{\Phi^{-1}(\lambda)\sqrt{1+\beta^2\sigma_u^2}}{\beta}\right)^2\right]\right\};$$

Note that as β goes to infinity, the criterion function goes to

$$-E_\theta\left\{\frac{1}{t}+\frac{(\bar{z}-\gamma-\Phi^{-1}(\lambda))^2}{s}\right\}.$$

In this case, a proper prior for β will ensure that the expectation is finite so that the maximization process can proceed.

The situation is worse when $B(\theta)$ is the identity matrix since

$$tr(B(\theta)I(\theta,\eta)^{-1}) = O(\beta^6).$$

The prior of β should decay to zero at least as fast as $o(\beta^{7})$ to guarantee that the expectation is finite.

5 Numerical Investigation

Numerical investigations are performed to illustrate the effect of meausurement error on the design of the following model

$$p(y=1 \mid x) = \Phi(\beta(x-\gamma)).$$

σ_u^2 is set to 0.26 and to 0.52. γ is assumed to be uniformly distributed over (0,2). Three proper prior distributions for β are considered. The first is the uniform distribution over the range (0,2). The other 2 priors are suggested by Tang (1992). Since all the three priors are non-zero at $\beta=0$, only ϕ_1 and ϕ_3 designs are studied. The design measure for the ϕ_1 and ϕ_3 optimal designs using the three proper prior distributions for β are found as follows.

Table 1. ϕ_1 and ϕ_3 optimal designs for probit regression with errors-in-variables

		$\gamma \sim U(0,2)$, $\beta \sim U(0,2)$		$\gamma \sim U(0,2)$, $\beta_r \sim U(0,\sigma_u^{-1})$		$\gamma \sim U(0,2)$, $\beta \sim \exp(-\sigma_u\beta)$	
		Point	Weight	Point	Weight	Point	Weight
$\sigma_u^2 = .26$	ϕ_1	2.50	0.5	2.74	0.5	2.73	0.5
		-0.48	0.5	-0.75	0.5	-0.74	0.5
	ϕ_3	2.29	0.5	2.37	0.5	2.30	0.5
		-0.29	0.5	-0.37	0.5	-0.30	0.5
$\sigma_u^2 = .52$	ϕ_1	2.94	0.5	3.37	0.5	3.36	0.5
		-0.92	0.5	-1.39	0.5	-1.37	0.5
	ϕ_3	2.57	0.5	2.89	0.5	2.79	0.5
		-0.58	0.5	-0.90	0.5	-0.79	0.5

It is found that a two point design is enough for all the situations, i.e. Whittle's condition is met, and the design points are symmetrically distributed about γ. As the variance of the distribution of measurement error increases, the deviation of the design points from γ also increases. Furthermore, the design points for the ϕ_3 design seem to be closer to γ than that for the ϕ_1 design.

When no measurement error is present and assuming the priors for both γ and β are uniformly distributed over (0,2), the ϕ_1 and ϕ_3 optimal designs are found as follows:

Table 2. ϕ_1 and ϕ_3 optimal designs for probit regression without errors-in-variables

	Point	Weight
ϕ_1	1.93	0.5
	0.08	0.5
ϕ_3	1.93	0.5
	0.07	0.5

Let $\phi(\eta)$ be the criterion function value for the ϕ optimal design at the measure η. We define the efficiency, Eff, in terms of equivalent sample size, of an ϕ_1 optimal design adjusted for measurement error as

$$Eff = \sqrt{\exp\ (\phi_1(\eta_1) - \phi_1(\eta_0)\ )}\ .$$

Similarly, for ϕ_3, the efficiency is defined as

$$Eff = \frac{\phi_3(\eta_1)}{\phi_3(\eta_0)}$$

where ϕ_i, (i=1,3), is the optimal design adjusted for measurement error and η_0 is the corresponding measure whereas η_1 is the measure that assumes no measurement error. For the numerical example, we determine the efficiencies of the ϕ_1 and ϕ_3 optimal designs using uniform priors over (0,2) for both γ and β. The results are as follows:

Table 3. Efficiencies of ϕ_1 and ϕ_3 optimal designs.

	Efficiency for ϕ_1 (%)	Efficiency for ϕ_3 (%)
$\sigma_u^2 = 0.26$	91.5	92.6
$\sigma_u^2 = 0.52$	83.6	83.3

It can be seen that the ϕ_1 design is more affected by small measurement error than the ϕ_3 design is. But ϕ_3 design is more affected by large measurement error than the ϕ_1 design is. Moreover, the efficiency declines as the variability of the measurement error distribution increases.

Another numerical investigation is performed using a standard gamma distribution, with shape parameter 3, as prior for β and a uniform prior, over the range (0,2), for γ. Note that the prior for β is zero at $\beta=0$ and decays to zero at an exponential rate in the tail so that all the three designs can now be evaluated. ϕ_2 design is taken to be the one that γ is the only concern. The results are as follows.

Table 4. ϕ_1, ϕ_2 and ϕ_3 optimal designs for probit regression with errors-in-variables

	ϕ_1		ϕ_2		ϕ_3	
	Point	Weight	Point	Weight	Point	Weight
$\sigma_u^2 = .26$	-0.08	0.5	0.15	0.5	0.10	0.5
	2.08	0.5	1.85	0.5	1.90	0.5
$\sigma_u^2 = .52$	-0.40	0.5	0.07	0.5	-0.16	0.5
	2.40	0.5	1.93	0.5	2.16	0.5

Again, a two point design is sufficient for all the criteria and the design points are further away from the prior mean for γ as the variance of the measurement error distribution increases. It can be seen that the design points for the ϕ_3 design lie between that of the ϕ_1 and the ϕ_2 designs. For the case of ϕ_2 optimal design, the design points cannot be found using priors over the half real line when there is no measurement error (Chaloner and Larntz, 1989). On the other hand, the design points can be determined when there is measurement error.

6 Conclusion

The aim of the study of an optimal design is to design an experiment which provides as precise as possible the information about the point estimates of the regression model under consideration. A Bayesian optimal design is one such that the pre-posterior loss is minimized. The A-optimal, ϕ_1, and D-optimal, ϕ_2, designs are appropriate when the posterior mean is used as the point estimate. However, because of the inadmissible property of the posterior mean for binary regression models with Berkson type measurement error (Tang and Bacon-Shone, 1992), both the aforementioned designs are not appropriate as long as there is measurement error in the covariate. Since Tang and Bacon-Shone suggest to use the posterior mode, it seems natural to devise a corresponding design, ϕ_3. Moreover, this research appears to be the first study of an optimal design using the posterior mode as the point estimate. From the simulation study, the ϕ_3 design is found to be a useful alternative to the well known ϕ_1 design as it is more efficient. To summarize, for a probit model with measurement error in the covariate, a proper prior for the slope parameter is needed, the posterior mode should be used as the Bayes point estimate and an optimal design should employ the ϕ_3 criterion.

Reference

Abdelbasit, K.M. and Plackett, R.L. (1983). Experimental Design for Binary Data. Journal of the American Statistical Association, 78, 90-98.

Burr, D. (1988). On Errors-in-variables in Binary Regression - Berkson's Case. Journal of the American Statistical Association, 83, 739-743.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian Design Applied to Logistic Regression Experiments. Journal of Statistical Planning and Inference, 21, 191-208.

Efron, B.F. and Hinkley, D.V. (1978). Accessing the Accuracy of the Maximum Likelihood Estimator: Observed versus Expected Fisher Information. Biometrika, 65, 457-487.

Graybill, F.A. (1983). Matrices with Applications in Statistics, Second Edition. Wadsworth International Group, Belmont, California.

Magnus, J.R. and Neudecker H. (1988). Matrix Differential Calculus with Applications in Statistics and Econometrics. John Wiley & Sons.

Minkin, S. (1987). Optimal designs for Binary Data. Journal of the American Statistical Association, 82, 1098-1103.

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics. Wiley, New York.

Tang, P.K. (1992). Bayesian Analysis of Errors-in-variables in Generalized Linear Models. Ph.D. thesis, University of Hong Kong.

Tang, P.K. and Bacon-Shone, J. (1992). Bayesian Analysis of Errors-in-variables in Binary Regression Models. Submitted for publication.

Appendix

We need the following lemma to prove Theorem 1.

Lemma

Let η_1 and η_2 be two different design measures and let π be a real number such that $0 \le \pi \le 1$, we have

$$I(\theta, \pi\eta_1 + (1-\pi)\eta_2) = \pi I(\theta, \eta_1) + (1-\pi) I(\theta, \eta_2)$$

The proof of the lemma is straight forward and omitted.

Theorem 1

For two distinct design measures, η_1 and η_2, and π, where $0 \le \pi \le 1$, we have

$$\sqrt{|I(\theta, \pi\eta_1 + (1-\pi)\eta_2)|}$$

$$\ge \pi\sqrt{|I(\theta, \eta_1)|} + (1-\pi)\sqrt{|I(\theta, \eta_2)|}$$

Proof

The Minkowski inequality for determinants of positive definite matrices (Magnus and Neudecker, 1988, p.227) stated that if A and B are $n \times n$ positive definite matrices, then

$$|A + B|^{\frac{1}{n}} \ge |A|^{\frac{1}{n}} + |B|^{\frac{1}{n}} .$$

Given that $I(\theta,\eta)$ is 2×2 and using the above lemma and the Minkowski inequality, we have

$$\sqrt{|I(\theta, \pi\eta_1 + (1-\pi)\eta_2)|}$$

$$= |\pi I(\theta, \eta_1) + (1-\pi) I(\theta, \eta_2)|^{\frac{1}{2}}$$

$$\ge |\pi I(\theta, \eta_1)|^{\frac{1}{2}} + |(1-\pi) I(\theta, \eta_2)|^{\frac{1}{2}}$$

$$= \pi |I(\theta, \eta_1)|^{\frac{1}{2}} + (1-\pi) |I(\theta, \eta_2)|^{\frac{1}{2}}$$

Hence the theorem is proved.

In fact, the theorem implies that the ϕ_3 criterion is concave since the inequality still holds when expectations w.r.t. $\hat{\theta}$ are taken on both sides of (3). The distribution of $\hat{\theta}$ can be approximated by using the prior distribution of θ as the predictive distribution of $\hat{\theta}$. The fact that the design criterion is concave allows us to apply Whittle's results for the determination of optimal design points. Hence we need to find the directional derivative

$$F(\eta,\eta_z) = \lim_{\epsilon\to\infty} \frac{\sqrt{|I(\theta,(1-\epsilon)\eta + \epsilon\eta_z)|} - \sqrt{|I(\theta,\eta)|}}{\epsilon} .$$

Let X, Y and Z be positive definite matrices. Silvey (1980) showed that

$$F(X , Y) = G(X , Y - X) .$$

where

$$G(X,Z) = \lim_{\epsilon\to\infty} \frac{\sqrt{|X + \epsilon Z|} - \sqrt{|X|}}{\epsilon} .$$

To find the directional derivative for the current design, we start with considering

$$\sqrt{|X + \epsilon Y|} - \sqrt{|X|}$$

$$= \sqrt{|X|\,|I + \epsilon X^{-1} Y|} - \sqrt{|X|}$$

$$= |X|^{\frac{1}{2}} \sqrt{|I + \epsilon X^{-1} Y|} - \sqrt{|X|}$$

Recall Theorem 8.3.4 of Graybill (1983) that if the k×k matrix C is given by

$$C = D + \alpha a b'$$

where D is a nonsingular diagonal matrix, a and b are each k×1 vectors, and α is a scalar such that

$$\alpha \neq -\left[\sum_{i=1}^{k} a_i b_i / d_{ii}\right]^{-1}$$

then

$$|C| = \left[1 + \alpha \sum_j \frac{a_j b_j}{d_{jj}}\right] \prod_i d_{ii} .$$

Therefore, we have

$$|I + \epsilon X^{-1} Y| = 1 + \epsilon\, tr\,(X^{-1} Y)$$

Then

$$\sqrt{|X + \epsilon Y|} - \sqrt{|X|}$$

$$= |X|^{\frac{1}{2}} |I + \epsilon(X^{-1}Y)|^{\frac{1}{2}} - |X|^{\frac{1}{2}}$$

$$= |X|^{\frac{1}{2}} \left(1 + \frac{\epsilon}{2} tr(X^{-1}Y) + O(\epsilon^2) \right) - |X|^{\frac{1}{2}}$$

$$= \frac{\epsilon}{2} |X|^{\frac{1}{2}} tr(X^{-1}Y) + O(\epsilon^2)$$

The directional derivative, G, is

$$G(X,Y) = \lim_{\epsilon \to \infty} \frac{\sqrt{|X + \epsilon Y|} - \sqrt{|X|}}{\epsilon}$$

$$= \frac{|X|^{\frac{1}{2}}}{2} tr(X^{-1}Y)$$

Hence, the directional derivative for the ϕ_3 design is

$$F(X , Y) = \frac{|X|^{\frac{1}{2}}}{2} tr(X^{-1}(Y - X))$$

$$= \frac{|X|^{\frac{1}{2}}}{2} tr(X^{-1}Y - I)$$

$$= \frac{|X|^{\frac{1}{2}}}{2} [tr(X^{-1}Y) - 2] .$$