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, Hongkong Bank, 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 \( \theta^T = (\theta_1, \ldots, \theta_d) \). By expanding the definition of a design to include any probability measure \( \eta \) on \( Z \) we define the normalized matrix \( I(\theta, \eta) \) by

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

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 $\eta_1$ and $\eta_2$ in $\mathcal{H}$ the derivative at $\eta_1$ in the direction of $\eta_2$ is defined by

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

The special case that $\eta_2$ 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 $\eta_0$ 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 $\mathcal{H}$, where $\mathcal{H}$ is the set of all probability measures on $X$, for which $\phi$ is finite; and
4. $\phi$ is such that if $\eta_n \to \eta$ in weak convergence then $\phi(\eta_n) \to \phi(\eta)$.

Note that $X$ in the assumptions refers to the experimental region. Under the Berkson error model, the proxy variable, $x$, 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 $\theta$, 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 $\eta$ for which $I(\theta, \eta)$ is singular for $\theta$ values of non-zero prior probability we define $\phi_1$ 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

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

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

The corresponding directional derivatives for the two criteria are
\[ d(\eta, z) = E_0 \text{tr} \left( I(\theta, z) I(\theta, \eta)^{-1} \right) - p ; \]

and

\[ d(\eta, z) = E_0 \text{tr} \left( B(\theta) I(\theta, \eta)^{-1} I(\theta, z) I(\theta, \eta)^{-1} \right) + \phi_2(\eta) . \]

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 \( \theta \) using a design measure \( \eta \) is approximately distributed as

\[ N_p (\hat{\theta}, nI(\hat{\theta}, \eta)^{-1}) \]

where \( \hat{\theta} \) is the maximum likelihood estimate of \( \theta \).

2 Probit Model

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

\[
p(y = 1 \mid z) = \int \Phi(\beta(z - \gamma)) \phi\left(\frac{x - z}{\sigma_u}\right) \, dx
\]

\[
= \Phi\left(\frac{\beta}{\sqrt{1 + \beta^2 \sigma_u^2}}(z - \gamma)\right)
\]

\[
= \Phi(\beta_s(z - \gamma))
\]
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 \(\sigma_u^2\) is known.

Here, \(\theta = (\gamma, \beta)\). The likelihood for \(\gamma\) and \(\beta\) is then
\[
L = \prod_{i=1}^{n} \Phi \left( \frac{\beta (z_i - \gamma)}{\sqrt{1 + \beta^2 \sigma_u^2}} \right)^{y_i} \left[ 1 - \Phi \left( \frac{\beta (z_i - \gamma)}{\sqrt{1 + \beta^2 \sigma_u^2}} \right) \right]^{1-y_i}.
\]

The expected Fisher's information matrix for \(\gamma\) and \(\beta\) is
\[
I = \begin{bmatrix}
\sum_{i=1}^{n} \frac{1}{\Phi_i (1 - \Phi_i)} \left( \frac{\partial \Phi_i}{\partial \gamma} \right)^2 & \sum_{i=1}^{n} \frac{1}{\Phi_i (1 - \Phi_i)} \left( \frac{\partial \Phi_i}{\partial \gamma} \right) \left( \frac{\partial \Phi_i}{\partial \beta} \right) \\
\sum_{i=1}^{n} \frac{1}{\Phi_i (1 - \Phi_i)} \left( \frac{\partial \Phi_i}{\partial \beta} \right) \left( \frac{\partial \Phi_i}{\partial \gamma} \right) & \sum_{i=1}^{n} \frac{1}{\Phi_i (1 - \Phi_i)} \left( \frac{\partial \Phi_i}{\partial \beta} \right)^2
\end{bmatrix}
\]

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
\[
\frac{\partial \Phi_i}{\partial \beta} = \frac{z_i - \gamma}{\left(1 + \beta^2 \sigma_u^2 \right)^{3/2}} \Phi \left( \frac{\beta (z_i - \gamma)}{\sqrt{1 + \beta^2 \sigma_u^2}} \right).
\]
3 Classical D-optimal Design

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

\[
| I | = \frac{n^2 \beta^2 c^2}{\Phi^2(1 - \Phi)^2} \Phi^4(\beta c),
\]

(1)

where

\[
\Phi = \Phi(\beta c).
\]

In the current situation, the classical approach would find \( c \) that maximizes (1) with \( \beta \) replaced by \( \beta_c \).

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

\[
\Phi^{-1}(p^*) = \frac{\beta}{\sqrt{1 + \beta^2 \sigma_w^2}}(z - \gamma)
\]

i.e.

\[
z = \frac{\Phi^{-1}(p^*) \sqrt{1 + \beta^2 \sigma_w^2}}{\beta} + \gamma.
\]

(2)

It can be seen that \( z \) is an increasing function in \( \sigma_w \) and a decreasing function in \( \beta \). Hence, we need to put the support points further away from \( \gamma \) if error is present.

Given a known value of \( \beta, \beta_c, c \) is determined from (1). Then
\[ p = \Phi(\beta, (z - \gamma)) \]

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

(3)

Consider the behaviour of \( p \) when \( \beta \to \infty \):

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

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

It can be seen that the conditional density does not go to 0 or 1 as \( \beta \) 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 \( \beta \).

4 Bayesian Optimal Designs

We consider Bayesian optimal designs under 3 criteria. The first criterion, \( \phi_1 \), is the D-optimal criterion, which is same as the Fisher’s information criterion. The second criterion, \( \phi_2 \), is the A-optimal criterion which is derived from using the posterior mean as the Bayes point estimate. The third criterion, \( \phi_3 \), 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, \( \eta \), putting \( \eta \) weight at \( k \) distinct design points \( z_i, i = 1, \ldots, k, \sum \eta_i = 1 \), we define
\[
\phi_i = \phi \left( \frac{\beta}{\sqrt{1 + \beta^2 \sigma_u^2}} (z_i - \gamma) \right) ; \quad \Phi_i = \Phi \left( \frac{\beta}{\sqrt{1 + \beta^2 \sigma_u^2}} (z_i - \gamma) \right) \\
\]

\[
w_i = \frac{\Phi_i^2}{\Phi_i (1 - \Phi_i)} ; \quad t = \sum_{i=1}^{k} n_i w_i \\
\]

\[
\bar{z} = \frac{\sum_{i=1}^{k} n_i w_i z_i}{\sum_{i=1}^{k} n_i w_i} ; \quad s = \sum_{i=1}^{k} n_i w_i (z_i - \bar{z})^2 .
\]

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

\[
I(\theta, \eta) = 
\frac{\beta t}{1 + \beta^2 \sigma_u^2} - \frac{\beta t (\bar{z} - \gamma)}{(1 + \beta^2 \sigma_u^2)^2} \\
\frac{\beta t (\bar{z} - \gamma)}{(1 + \beta^2 \sigma_u^2)^2} - \frac{s + \beta t (\bar{z} - \gamma)^2}{(1 + \beta^2 \sigma_u^2)^3}
\]

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

\[
I^{-1}(\theta, \eta) = 
\begin{pmatrix}
\frac{1}{t} & \frac{(\bar{z} - \gamma)^2 (1 + \beta^2 \sigma_u^2)}{\beta^2} & \frac{(\bar{z} - \gamma)(1 + \beta^2 \sigma_u^2)^2}{s \beta} \\
\frac{(\bar{z} - \gamma)(1 + \beta^2 \sigma_u^2)^2}{s \beta} & \frac{(1 + \beta^2 \sigma_u^2)^3}{s}
\end{pmatrix}
\]

4.1 \( \phi_i \) Optimal Design

It can be shown that the criterion function and the directional derivative for the \( \phi_i \) optimal design are respectively
\[
\phi_1(\eta) = \mathcal{E}_0 \left[ \log \left( \frac{\beta^2 t s}{(1 + \beta^2 \sigma_u^2)^4} \right) \right];
\]

and

\[
d(\eta, z) = \mathcal{E}_0 \left[ w \left( \frac{1}{t} + \frac{(z - \gamma)^2}{s} \right) \right] - 2.
\]

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 \(\phi_2\) Optimal Design

In the case of \(\phi_2\) 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 \(\tilde{z}_0\) 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 \(\phi_2\) optimality for the percentiles response point to be

\[
\phi_2(\eta) = -\mathcal{E}_0 \left\{ \frac{1 + \beta^2 \sigma_u^2}{\beta^2} \left[ \frac{1}{t} + \frac{1}{s} \left( \frac{\Phi^{-1}(\lambda) \sqrt{1 + \beta^2 \sigma_u^2}}{\beta} \right)^2 \right] \right\};
\]

and

\[
d(\eta, z) = \mathcal{E}_0 \left[ w \left( \frac{1}{t} + \frac{(z - \gamma)^2}{s} \right) \right] - 2.
\]
\begin{equation}
d(\eta, z) = R_0 \left\{ \frac{w}{\beta^2} \left[ \sqrt{1 + \beta^2 \sigma^2} \left( \frac{1}{t} + \frac{(z - y)(z - z)}{s} \right) - \Phi^{-1}(\lambda)(z - z)(1 + \beta^2 \sigma^2) \right] \right\}^2 + \phi_2(\eta)
\end{equation}

respectively.

Note that if we want to estimate \( y \) alone, we can set \( \lambda = 0.5 \). If both \( y \) and \( \beta \) 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 \( \phi_2 \), optimal design for the sum of parameters to be

\begin{equation}
\phi_2(\eta) = -R_0 \left\{ \left( \frac{1}{t} + \frac{(z - y)^2}{s} \right) \left( 1 + \beta^2 \sigma^2 \right) + \frac{(1 + \beta^3 \sigma^2)^3}{s} \right\}
\end{equation}

and

\begin{equation}
d(\eta, z) = \frac{w(1 + \beta^2 \sigma^2)}{\beta^2} \left( \frac{1}{t} + \frac{(z - y)(z - z)}{s} \right)^2 + \frac{w(z - z)(1 + \beta^2 \sigma^2)^3}{s^2}
\end{equation}

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

4.3 \( \phi_3 \) Optimal Design

If \( \bar{\theta} \) is the posterior mode, then, as the sample size gets large,

\begin{equation}
\bar{\theta} \sim N(\theta, [I(\bar{\theta})]^{-1})
\end{equation}

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

\begin{equation}
I_{ij} = -n \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(z | \theta) \right\}
\end{equation}

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 \( \phi_3 \) 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 φ₁-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

\[ \phi_3(\eta) = E_0 \left[ \frac{\beta^2 z s}{(1 + \beta^2 \sigma_u^2)^2} \right]^{1\over 2} \]

and

\[ d(\eta, z) = E_0 \left\{ \frac{1}{2} \frac{1}{\beta t^2 s^2} \left[ w \left( \frac{1}{t} + \frac{(\bar{z} - z)^2}{s} \right) - 2 \right] \right\} \]

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 \( d(\eta, z) \to \text{constant} \). 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 \( \beta \to \infty \),

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

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

\[ \omega_i \to \frac{\Phi_i^2 \left( \frac{z_i - \gamma}{\sigma_u} \right)}{\Phi_i \left( \frac{z_i - \gamma}{\sigma_u} \right) \left( 1 - \Phi_i \left( \frac{z_i - \gamma}{\sigma_u} \right) \right)} ; \]

We can see that \( \omega_i \) tends to a constant as \( \beta \to \infty \). This behaviour of the \( \omega_i \)'s will make \( t, s \) and \( \bar{z} \) tend to constants as \( \beta \) goes to infinity.
The criterion function of the $\phi_1$ optimal design is

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

The argument of the log function goes to 0 as $\beta$ goes to infinity.

Recall that the design criterion of $\phi_c$ optimal design for the percentile response points is given by

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

Note that as $\beta$ goes to infinity, the criterion function goes to

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

In this case, a proper prior for $\beta$ 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 $\beta$ 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 measurement error on the design of the following model

$$
p(\gamma = 1 \mid x) = \Phi(\bar{\beta} (x - \gamma)).
$$

$\sigma^2$ is set to 0.26 and to 0.52. $\gamma$ is assumed to be uniformly distributed over (0,2). Three proper prior distributions for $\bar{\beta}$ 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 $\bar{\beta} = 0$, only $\phi_1$ and $\phi_c$ designs are studied. The design measure for the $\phi_1$ and $\phi_c$ optimal designs using the three proper prior distributions for $\bar{\beta}$ are found as follows.
Table 1. \( \phi_1 \) and \( \phi_2 \) optimal designs for probit regression with errors-in-variables

<table>
<thead>
<tr>
<th>( \sigma^2 = .26 )</th>
<th>( \gamma \sim U(0,2) )</th>
<th>( \gamma \sim U(0,2) )</th>
<th>( \gamma \sim U(0,2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta \sim U(0,2) )</td>
<td>( \beta_1 \sim U(0, \sigma^2) )</td>
<td>( \beta \sim \exp(-\sigma\beta) )</td>
<td>( \beta_1 \sim U(0, \sigma^2) )</td>
</tr>
<tr>
<td>( \phi_1 )</td>
<td>2.50</td>
<td>0.5</td>
<td>2.74</td>
</tr>
<tr>
<td></td>
<td>-0.48</td>
<td>0.5</td>
<td>-0.75</td>
</tr>
<tr>
<td>( \phi_2 )</td>
<td>2.29</td>
<td>0.5</td>
<td>2.37</td>
</tr>
<tr>
<td></td>
<td>-0.29</td>
<td>0.5</td>
<td>-0.37</td>
</tr>
</tbody>
</table>

| \( \sigma^2 = .52 \) | \( \phi_1 \) | 2.94 | 0.5 | 3.37 | 0.5 | 3.36 | 0.5 |
| | -0.92 | 0.5 | -1.39 | 0.5 | -1.37 | 0.5 |
| \( \phi_2 \) | 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 \( \gamma \). As the variance of the distribution of measurement error increases, the deviation of the design points from \( \gamma \) also increases. Furthermore, the design points for the \( \phi_2 \) design seem to be closer to \( \gamma \) than that for the \( \phi_1 \) design.

When no measurement error is present and assuming the priors for both \( \gamma \) and \( \beta \) are uniformly distributed over \((0,2)\), the \( \phi_1 \) and \( \phi_2 \) optimal designs are found as follows:

Table 2. \( \phi_1 \) and \( \phi_2 \) optimal designs for probit regression without errors-in-variables

<table>
<thead>
<tr>
<th>( \phi_1 )</th>
<th>Point</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi_2 )</td>
<td>1.93</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \phi_2 )</th>
<th>Point</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi_2 )</td>
<td>1.93</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Let \( \phi(\eta) \) be the criterion function value for the \( \phi \) optimal design at the measure \( \eta \). We define the efficiency, \( \text{Eff} \), in terms of equivalent sample size, of an \( \phi \), optimal design adjusted for measurement error as

\[
\text{Eff} = \sqrt{\exp \left( \phi_1(\eta_1) - \phi_1(\eta_0) \right)}.
\]

Similarly, for \( \phi_3 \), the efficiency is defined as

\[
\text{Eff} = \frac{\phi_3(\eta_1)}{\phi_3(\eta_0)}
\]

where \( \phi_i \), \((i=1,3)\), is the optimal design adjusted for measurement error and \( \eta_0 \) is the corresponding measure whereas \( \eta_1 \) is the measure that assumes no measurement error. For the numerical example, we determine the efficiencies of the \( \phi_1 \) and \( \phi_3 \) optimal designs using uniform priors over \((0,2)\) for both \( \gamma \) and \( \beta \). The results are as follows:

Table 3. Efficiencies of \( \phi_1 \) and \( \phi_3 \) optimal designs.

<table>
<thead>
<tr>
<th>( \sigma_\epsilon^2 )</th>
<th>Efficiency for ( \phi_1 ) (%)</th>
<th>Efficiency for ( \phi_3 ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.26</td>
<td>91.5</td>
<td>92.6</td>
</tr>
<tr>
<td>0.52</td>
<td>83.6</td>
<td>83.3</td>
</tr>
</tbody>
</table>

It can be seen that the \( \phi_1 \) design is more affected by small measurement error than the \( \phi_3 \) design is. But \( \phi_3 \) design is more affected by large measurement error than the \( \phi_1 \) 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 \( \beta \) and a uniform prior, over the range \((0,2)\), for \( \gamma \). Note that the prior for \( \beta \) 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. \( \phi_2 \) design is taken to be the one that \( \gamma \) is the only concern. The results are as follows.
Table 4. $\phi_1$, $\phi_2$, and $\phi_3$ optimal designs for probit regression with errors-in-variables

<table>
<thead>
<tr>
<th>$\phi_1$</th>
<th>Point</th>
<th>Weight</th>
<th>$\phi_2$</th>
<th>Point</th>
<th>Weight</th>
<th>$\phi_3$</th>
<th>Point</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_c^2 = .26$</td>
<td>-0.08</td>
<td>0.5</td>
<td>0.15</td>
<td>0.5</td>
<td>0.10</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.08</td>
<td>0.5</td>
<td>1.85</td>
<td>0.5</td>
<td>1.90</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_u^2 = .52$</td>
<td>-0.40</td>
<td>0.5</td>
<td>0.07</td>
<td>0.5</td>
<td>-0.16</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.40</td>
<td>0.5</td>
<td>1.93</td>
<td>0.5</td>
<td>2.16</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Again, a two point design is sufficient for all the criteria and the design points are further away from the prior mean for $\gamma$ as the variance of the measurement error distribution increases. It can be seen that the design points for the $\phi_3$ design lie between that of the $\phi_1$ and the $\phi_2$ designs. For the case of $\phi_3$ optimal design, the design points cannot be found using priors over the half real line when there is no measurement error (Chaloner and Larnitz, 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, $\phi_1$, and D-optimal, $\phi_2$, 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, $\phi_3$. 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 $\phi_3$ design is found to be a useful alternative to the well known $\phi_1$ 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 $\phi_3$ criterion.

Reference


Appendix

We need the following lemma to prove Theorem 1.

Lemma
Let $\eta_1$ and $\eta_2$ be two different design measures and let $\pi$ be a real number such that $0 \leq \pi \leq 1$, 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 straightforward and omitted.

Theorem 1
For two distinct design measures, $\eta_1$ and $\eta_2$, and $\pi$, where $0 \leq \pi \leq 1$, we have

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

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

$$||A + B||^\frac{1}{2} \geq ||A||^\frac{1}{2} + ||B||^\frac{1}{2}$$

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

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

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

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

Hence the theorem is proved.

In fact, the theorem implies that the $\phi_1$ 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 $\theta$ 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

\[
F(\eta_1, \eta_2) = \lim_{\epsilon \to \infty} \frac{\sqrt{I(\theta, (1 - \epsilon) \eta + \epsilon \eta_2)} - \sqrt{I(\theta, \eta)}}{\epsilon}.
\]

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

\[
\sqrt{X + \epsilon Y} - \sqrt{X} = \sqrt{X} - \sqrt{X} \frac{\epsilon Y}{\sqrt{X} + \sqrt{X} Y}.
\]

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

\[
C = D + \alpha a b^t,
\]

where \( D \) is a nonsingular diagonal matrix, \( a \) and \( b \) are each \( k \times 1 \) vectors, and \( \alpha \) is a scalar such that

\[
\alpha \neq -\sum_{i=1}^{k} \frac{a_i b_i}{d_{ii}}.
\]

then

\[
\begin{align*}
|C| &= \left|1 + \alpha \sum_{j=1}^{k} \frac{a_j b_j}{d_{jj}}\right| \prod_{i=1}^{k} d_{ii}.
\end{align*}
\]

Therefore, we have

\[
|I + \epsilon X^{-1} Y| = 1 + \epsilon \text{tr} (X^{-1} Y).
\]
Then

\[
\sqrt{ | X + \epsilon Y | } - \sqrt{ | X | } \\
= | X |^{\frac{1}{2}} | I + \epsilon ( X^{-1} Y ) |^{\frac{1}{2}} - | X |^{\frac{1}{2}} \\
= | X |^{\frac{1}{2}} \left( 1 + \frac{\epsilon}{2} \text{tr}(X^{-1} Y) + O(\epsilon^2) \right) - | X |^{\frac{1}{2}} \\
= \frac{\epsilon}{2} | X |^{\frac{1}{2}} \text{tr}(X^{-1} Y) + O(\epsilon^2)
\]

The directional derivative, \( G \), is

\[
G(X,Y) = \lim_{\epsilon \to 0} \frac{\sqrt{ | X + \epsilon Y | } - \sqrt{ | X | } }{\epsilon} \\
= | X |^{\frac{1}{2}} \text{tr}(X^{-1} Y)
\]

Hence, the directional derivative for the \( \phi \), design is

\[
F(X,Y) = \frac{1}{2} | X |^{\frac{1}{2}} \text{tr}(X^{-1}(Y - X)) \\
= \frac{1}{2} | X |^{\frac{1}{2}} \text{tr}(X^{-1} Y - I) \\
= \frac{1}{2} | X |^{\frac{1}{2}} \left[ \text{tr}(X^{-1} Y) - 2 \right]
\]

- 18 -