Journal of the Korean
OR/ MS Society
Vol. 20, No. 3, December 1995

A Bayesian Approach to Linear Calibration Design Problem*

Sung Chul Kim**

Abstract

Based on linear models, the inference about the true measurement x_f and the optimal designs x $(nx1)$ for the calibration experiment are considered via Bayesian statistical decision analysis. The posterior distribution of x_f given the observation $y_f(qx1)$ and the calibration experiment is obtained with normal priors for x_f and for the model parameters (α, β). This posterior distribution is not in the form of any known distributions, which leads to the use of a numerical integration or an approximation for the calculation of the overall expected loss. The general structure of the expected loss function is characterized in the form of a conjecture. A near-optimal design is obtained through the approximation of the conditional covariance matrix of the joint distribution of $(x_f, y_f^T)^T$. Numerical results for the univariate case are given to demonstrate the conjecture and to evaluate the approximation.

Key words: calibration, optimal design, linear model, Bayesian analysis, near optimal

1. INTRODUCTION

A measuring instrument must be calibrated for accurate inferences based on the relationship between two quantities. The calibration problem consists of two experiments: the future measuring experiment and the calibration experiment. The inference problem is to make inferences about an unknown true value x_f, from a random observation, q-vector y_f.

* This article is published with the support from 1994 Soongsil University Research Fund.
** Department of Statistics, Soongsil University, Seoul 156-743, Korea

The design problem is to find an optimal calibration experimental design vector x ($nx1$), which minimizes some overall loss function.

Hoadley [10] considers univariate ($q=1$) calibration and points out some difficulties with the maximum likelihood estimator for x_f. He justifies the inverse estimator as a posterior mean for x_f with a t prior. Brown [5] generalized Hoadley's results with a multivariate setup. Hunter and Lamboy [11] obtained a posterior distribution of x_f which has infinite variance. They discussed a controversy concerning infinite variances and inverse regression. Optimal design for linear regression model is discussed in Chaloner [7], in which the Bayesian optimal design for estimating linear combination of the regression parameters and a geometric interpretation of an optimal one-point design which parallels Elfving [9] are given. Brooks [4] discusses the problem of designing an experiment to help control the dependent variable at a prechosen value in a linear regression model.

Optimal calibration designs are studied in Barlow, Mensing and Smiriga [1] [2], where they use Bayesian approach with univariate formulation. Using bivariate normal prior for α and β, Barlow, Mensing and Smiriga [2] demonstrated that the expected loss function can be calculated by a computer program using three nested subroutines for each design. Buonaccorsi [6] used the classical estimator to minimize asymptotic variance.

The two experiments can be expressed as following linear model:

$$Y = 1\ \alpha^T + (x - x_0 1)\beta^T + E \tag{1.1}$$

$$Y_f = \alpha + (x_f - x_0)\beta + \varepsilon_f \tag{1.2}$$

where x_f is a scalar, α, β, y_f, and ε_f are q-vectors, 1 is an n-vector of 1's, and Y and E are (nxq) matrices. It is assumed that ε_i given x and is $N_q(0, \Gamma)$. It is also assumed that the errors are independent from observation to observation, i. e., ε_i is independent of ε_j given Γ for $i \neq j$. x_0 is the prior mean of x_f, which is the choice for the center of the model. The model with a single future experiment is sufficiently general since the multiple future experiments can be reduced to our model when the error variance Γ is known. The model with scalar independent variable is preferred in the sense that it is much more tractable, especially in the design problem. Another important consideration for scalar x_f is that if the number of response variables is less than the number of independent variables then x_f cannot be completely determined even when α and β are known (Brown [5]).

In addition to the independent normal error assumption, we assume that x_f is independent of (α, β, ,x, Y), and that y_f is independent of (x, Y) given (α, β). In other words, the future true value is independent of the parameters and the calibration experiment a priori, and the future observation is independent of the calibration experiment given the parameters. x_f is

independent of (x, Y) since the calibration experiment provides no information on x_f. The error covariance matrix, Γ, is assumed to be known. If Γ is unknown, we can use a vague prior or a natural conjugate prior for the parameters of a normal distribution as in DeGroot [8]. Furthermore, we assume that the loss function $\ell(d,x_f)$ depends only on x_f and d—the decision regarding x_f after observing y_f. For example, we do not consider the cost of the experiment in the loss function. It is also assumed that the set Ω of feasible experimental designs is bounded.

The data from the calibration experiment consist of (x_i, y_i), $i=1,2,\cdots,n$, where $x = (x_1,\cdots,x_n)^T$ are specified in advance. Based on our prior $p(x_f)$ and the model (1. 1) and (1. 2), our problem is to determine the experimental design x subject to some feasibility constraints so as to minimize overall expected loss. The posterior distribution of the parameters, $p(\alpha, \beta|x,Y)$, can be obtained by Bayes' Theorem, and the predictive distribution of y_f can be calculated as

$$p(y_f|x_f, x,Y) = \int\int p(y_f|\alpha, \beta, x_f)\, p(\alpha, \beta|x,Y)\,d\alpha\,d\beta. \tag{1. 3}$$

After the posterior distribution of x_f given (y_f,Y,x) has been obtained, the decision d should be made to minimize

$$E_{x_f,y_f|Y,x}[\ell(d,x_f)|y_f,Y,x].$$

If the loss function is the squared error loss, $(d-x_f)^2$, the decision d is the estimate of x_f after observing the future y_f. The best estimator is the posterior mean $d^* = E(x_f|y_f,Y,x)$, and the posterior risk after observing y_f is the posterior variance $Var(x_f|y_f,Y,x)$.

At the time of the decision regarding the experimental design, we do not know y_f or the test result Y. Therefore, the overall expected loss R(x) is the preposterior risk

$$R(x) \equiv E_{Y|x}E_{y_f|Y,x} \operatorname*{Min}_{d} E_{x_f|y_f,Y,x}[\ell(d,x_f)|y\ Y,x] \tag{1.4}$$

We must minimize R(x) with respect to $x = (x_1,\cdots,x_n)^T$. With the squared error loss, R(x) becomes

$$R(x) = E_{Y|x}E_{y_f|Y,x}\, Var(x_f|y_f,Y,x)\ .$$

In Section 2, we discuss upper and lower bounds on the overall risk function. They are based on the value of sample information and the value of perfect information, respectively. In Section 3, the general structure of R(x) is characterized based on the posterior distributions of (α, β) and of x_f. An optimal solution procedure is suggested in the form of a conjecture for the structure of optimal designs. In Section 4, a simple approximation is introduced by replacing the covariance matrix of $(x_f, y_f^T)^T$ given (α, β) by some constant matrix to obtain a near-optimal design. Numerical examples illustrating the conjecture and the approximation are given in Section 5.

2. BOUNDS ON THE RISK FUNCTION

Consider a series of calibration experiments and a future experiment. The quantity

$$E_{y_c} \underset{d}{Min}\ E_{x_f,y_f}[\ell(d,x_f)|y_f] - R(x) \tag{2.1}$$

is the expected value of sample information (EVSI, [13]) gained by performing the calibration experiment. The quantity

$$\underset{d}{Min}\ E_{x_f}[\ell(d,x_f)] - R(x) \tag{2.2}$$

is the EVSI gained by performing both experiments. Suppose, on the other hand, that we are given the true value of x_f at the time of decisions. Then the expected value of perfect information (EVPI, [13]) about x_f is

$$R(x) - E_{x_f} \underset{d}{Min}\ [\ell(d,x_f)] \tag{2.3}$$

If we can have the true value of (α, β) at the time of decision then the EVPI about (α, β) is

$$R(x) - E_{\alpha,\beta} E_{y_f|\alpha,\beta} \underset{d}{Min}\ E_{x_f|y_f,\alpha,\beta}\ [\ell(d,x_f)]\ y_f,\ \alpha,\beta] \tag{2.4}$$

It can be shown that the above EVSI's and EVPI's - the quantities (2.1) through (2.4) - are all nonnegative, that is,

$$LB_2 \le LB_1 \le R(x) \le UB_1 \le UB_2, \tag{2.5}$$

where

$$\begin{aligned}
R(x) &= E_{Y|x}\ E_{y_f|Y,x} \underset{d}{Min}\ E_{x_f|y_f,Y,x}\ [\ell(d,x_f)|y_f,\ Y,x], \\
UB_1 &= E_{y_f} \underset{d}{Min}\ E_{x_f|y_f}[\ell(d,x_f)|y_f], \\
UB_2 &= \underset{d}{Min}\ E_{x_f}[\ell(d,x_f)], \\
LB_2 &= E_{x_f} \underset{d}{Min}\ [\ell(d,x_f)], \qquad \text{and} \\
LB_1 &= E_{\alpha,\beta}\ E_{y_f|\alpha,\beta} \underset{d}{Min}\ E_{x_f|y_f,\alpha,\beta}\ [\ell(d,x_f)|y_f,\ \alpha,\beta].
\end{aligned}$$

Intuitively, the above inequalities states that the more information we have at the time of decision, the smaller the final expected loss will be. We can use either UB_1 or UB_2 as an upper bound for R(x). LB_2 is the expected loss when we have perfect information about x_f, and LB_1 is the expected loss if we have perfect information about (α, β) before the future experiment. If we let

$$LB_3 = E_{Y|x}\, E_{\alpha,\beta|Y,x}\, E_{y_f|\alpha,\beta}\, \underset{d}{Min}\, E_{x_f|y_f,\alpha,\beta}\,[\,\ell(d,x_f)|y_f,\ \alpha,\ \beta], \tag{2.6}$$

then LB_3 is the expected loss if we have perfect information about (α, β) after the calibration experiment but before the future experiment. Kim [12] has shown that $LB_1 = LB_3$, and this fact verifies that the purpose of the calibration experiment is to learn about (α, β). Any of these LB's can be used as a lower blund for R(x) depending on the situation. Upper and lower bounds are useful in checking computer calculation in numerical examples. Details of the above argument can be found in Kim [12].

When $\ell(d,x_f)=(d-x_f)^2$, the above inequalities (2. 5) become

LEMMA 2.1. *If the range of possible decisions, d does not depend on* x *or* Y *and if* $\ell(d,x_f)=(d-x_f)^2$, *then we have*

$$\begin{aligned} 0 &\le E_{\alpha,\beta}\, E_{y_f|\alpha,\beta} Var(x_f|y_f,\ \alpha,\beta) \\ &\le E_{Y|x}\, E_{y_f|Y,x}\, Var(x_f|y_f,\ Y,x) = R(x) \\ &\le E_{y_f}\, Var(x_f|y_f) \\ &\le Var(x_f). \end{aligned}$$

If, in addition, x_f *is* $N(x_0,\ \sigma_0^2)$ a priori, then the second term on the first line becomes

$$E_\beta[(1/\sigma_0^2+\beta^T\ \Gamma^{-1}\ \beta)^{-1}], \tag{2.7}$$

which requires q-multiple integration to evaluate. The fourth term (third line) can be calculated in $q+1$ integrals if $(\alpha^T,\ \beta^T)^T$ is normal. In this case, it is easy to see, without any integration, that y_f given x_f is normal. And $p(y_f)$, $p(x_f|y_f)$ and $Var(x_f|y_f)$ can be computed in one integration. It takes q-multiple integration to calculate

$$E_{y_f}\, Var(x_f|y_f) = \int p(y_f)\, Var(x_f|y_f)dy_f.$$

It turns out that R(x) needs $(2q+1)$-integration to evaluate, so the computational time for these bounds, N^q (or N^{q+1}), is much less than N^{2q+1} for R(x), where N is the number of grid points in a numerical integration. A little less favorable lower bound than (2.7) can be found easily with Jensen's inequality so that for univariate case,

$$R(x) \ge [1/\sigma_0^2+(b^2+\sigma_b^2)/\sigma^2]^{-1}, \tag{2.8}$$

where b is the prior mean of β, σ_b^2 is the prior variance of β, and σ^2 is the variance of the error ε.

It follows from (2. 5) that the expected loss function can only decrease if we perform additional calibration experiments. This may not be true if $\ell(\ ,\cdot\ )$ were to depend on (x,Y).

LEMMA 2.2. *If the range of possible decisions, d, does not depend on* x *or* Y, *we have*

$$R(x_1,\cdots,x_n) \geq R(x_1,\cdots,x_n,x_{n+1}),$$

where the first n coordinates are the same on both sides of inequality.

3. STRUCTURE OF THE OPTIMAL EXPERIMENTAL DESIGN

3.1. Likelihood and Posterior Distributions

The likelihood for (α,β) given the data is

$$L(\alpha,\beta|data,\ x_0) \propto \exp\ [-1/2tr\ \Gamma^{-1}\ [Y-1\alpha^T-(x-x_0 1)\beta^T]^T\ [Y-1\alpha^T-(x-x_0 1)\beta^T]]. \tag{3.1}$$

The likelihood depends on n, $\sum_{i=1}^{n}(x_i-x_0)$, $\sum_{i=1}^{n}(x_i-x_0)^2$, Y^T1, and $Y^T(x-x_0 1)$. Assume a priori $E(\alpha) = a$, $E(\beta) = b$. Letting $U = Y-1a^T-(x-x_0 1)b^T$, we can see that the posterior distribution of (α, β) depends on the data only through n, $\sum_{i=1}^{n}(x_i-x_0)$, $\sum_{i=1}^{n}(x_i-x_0)^2$, z_1, and z_2, where $z_1 = U^T1$, and $z_2 = U^T(x-x_0 1)$ are the results of the calibration experiment. It follows that

$$z_1 \mid \alpha,\beta,x \sim N_q\ (\mu_1,\ n\Gamma),$$

$$z_2 \mid \alpha,\beta,x \sim N_q\ [\mu_2,\ \sum_{i=1}^{n}(x_i-x_0)^2\Gamma],$$

and the covariance matrix

$$Cov(z_1,\ z_2|\alpha,\beta,x) = E[(\sum_{i=1}^{n}\varepsilon_i)[\sum_{i=1}^{n}(x_i-x_0)\ \varepsilon_i^T)]] = \sum_{i=1}^{n}(x_i-x_0)\ \Gamma,$$

where

$$\mu_1 = n(\alpha-a)+\sum_{i=1}^{n}(x_i-x_0)(\beta-b),$$

and

$$\mu_2 = \sum_{i=1}^{n}(x_i-x_0)(\alpha-a)+\sum_{i=1}^{n}(x_i-x_0)^2(\beta-b).$$

Jointly, $\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ given α,β and x is

$$N_{2q}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix},\ \begin{bmatrix} n\Gamma & \sum(x_i-x_0)\Gamma \\ \sum(x_i-x_0)\Gamma & \sum(x_i-x_i)^2\Gamma \end{bmatrix}\right)$$

which depends on x only through n, $\sum_{i=1}^{n}(x_i-x_0)$, $\sum_{i=1}^{n}(x_i-x_0)^2$. Lemma 3.1 summarizes the above result, which is a multivariate version of Barlow, Mensing and Smiriga [1].

LEMMA 3.1. *For all loss functions* $\ell(d,x_f)$ *of* d *and* x_f *and all priors on* (α, β) *and* x_f, R(x) *depends on* x *only through* n, $(\bar{x}-x_0)=\sum_{i=1}^{n}(x_i-x_0)/n$ *and* $s_x^2=\sum_{i=1}^{n}(x_i-x_0)^2/n$.

Suppose a priori that (α, β) is independent of ε. With a normal prior (3. 2) for (α, β)

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix} \sim N_{2q}\left[\begin{bmatrix}a\\ b\end{bmatrix},\begin{bmatrix}D_a & 0\\ 0 & D_b\end{bmatrix}\right], \tag{3.2}$$

where D_a and D_b are diagonal matrices, the posterior distribution of (α,β) given x and Y is

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\Bigg| x,Y,\ x_0 \sim N_{2q}\left[\begin{bmatrix}\mu_\alpha\\ \mu_\beta\end{bmatrix},\ S^{-1}=\begin{bmatrix}S_{11} & S_{12}\\ S_{21} & S_{22}\end{bmatrix}^{-1}\right], \tag{3.3}$$

where

$$S_{11}= D_a^{-1}+n\Gamma^{-1}$$

$$S_{22}= D_b^{-1}+\sum_{i=1}^{n}(x_i-x_0)^2\Gamma^{-1}$$

$$S_{12}=S_{21}= \sum_{i=1}^{n}(x_i-x_0)\Gamma^{-1}$$

and

$$\mu_\alpha = a+\left\{C_1\ C_2-[\sum_{i=1}^{n}(x_i-x_0)]^2 I\right\}^{-1}\left\{C_1\ z_1-[\sum_{i=1}^{n}(x_i-x_0)]z_2\right\}$$

$$\mu_\beta = b+\left\{C_2\ C_1-[\sum_{i=1}^{n}(x_i-x_0)]^2 I\right\}^{-1}\left\{C_2\ z_2-[\sum_{i=1}^{n}(x_i-x_0)]z_1\right\},$$

where

$$C_1 = \sum_{i=1}^{n}(x_i-x_0)^2 I+\Gamma D_b^{-1}$$

$$C_2 = nI+\Gamma D_a^{-1}$$

As expected with a normal likelihood model, the posterior covariance matrix, S^{-1}, does not depend on the test results Y—here z_1 and z_2.

Recall the future experiment model

$$y_f = \alpha + (x_f-x_0)\beta+ \varepsilon_f. \tag{1.2}$$

Since α and β are jointly normal given (x_f, Y, x), it follows that $\alpha+(x_f-x_0)\beta=A\begin{bmatrix}\alpha\\ \beta\end{bmatrix}$ given (x_f, Y, x), is normal with mean $A\begin{bmatrix}\mu_\alpha\\ \mu_\beta\end{bmatrix} = \mu_\gamma + (x_f-x_0)\mu_\beta$ and covariance matrix $AS^{-1}A^T$, where $A=[I\ (x_f-x_0)I]$. Therefore, the predictive distribution of y_f is

$$y_f|x_f, Y, x \sim N_q\ [\ \mu_\alpha+(x_f-x_0)\mu_\beta,\ AS^{-1}A^T+\Gamma], \tag{3.4}$$

since (α, β) is independent of ε given (x_f, Y, x). For the prior distribution of x_f, we suppose $x_f \sim N(x_0, \sigma_0^2)$. Then by Bayes' Theorem, we have the posterior distribution of x_f

$$\begin{aligned} p(x_f|y_f, Y, x) &\propto p(y_f|x_f, Y, x)p(x_f) \\ &\propto |2\pi(AS^{-1}A^T+\Gamma)|^{-1/2} \exp\{-1/2[(x_f-x_0)^2/\sigma_0^2+[y_f-\mu_\alpha-(x_f-x_0)\mu_\beta]^T \\ &\quad (AS^{-1}A^T+\Gamma)^{-1}[y_f-\mu_\alpha-(x_f-x_0)\mu_\beta]\} \end{aligned} \tag{3.5}$$

As $A = [I\ (x_f-x_0)I]$ depends on x_f, this distribution cannot be a normal distribution. This non-normal posterior distribution of x_f makes the problem more difficult to analyze, and we need a numerical calculation for either the decision regarding x_f or the calibration design x.

3.2. Symmetry of R(x)

The calibration experiment should be designed so that the expected value of the loss function is minimized. A design is an n-vector x, but we can reduce the dimension of the design to three, namely n, $\bar{x}-x_0 = \sum_{i=1}^{n}(x_i-x_0)/n$ and $s_x^2 = \sum_{i=1}^{n}(x_i-x_0)^2/n$ by using Lemma 3.1. If our design space Ω in terms of x is, R^n, this transformation of the design space gives a constraint on the new design space Ξ_n, that is, since

$$\sum_{i=1}^{n}(x_i-\bar{x})^2/n \geq 0,$$

it follows that

$$\sum_{i=1}^{n}(x_i-x_0-\bar{x}+x_0)^2/n = \sum_{i=1}^{n}(x_i-x_0)^2/n-(\bar{x}-x_0)^2 \geq 0,$$

and hence

$$|\bar{x}-x_0| \leq s_x.$$

Notice that this transformation is valid only for $n \geq 2$ because if $n = 1$ then $(\bar{x}-x_0)^2$ is identical to s_x^2.

Since $\bar{x}-x_0$ and s_x are symmetric functions of an experimental design x, it follows that,

for fixed n, any permutation of coordinates of an experimental design solution is also a solution (unless violates the feasibility constraints). From here on, we will use the design x to represent either $x = (x_1, \cdots, x_n)^T$ or equivalently $x = (n, \bar{x} - x_0, s_x)^T$, and make no distinction between them.

Consider an n-vector x^* which is optimal when restricted to n-vectors, and an (n+1)-vector w^* which is optimal to (n+1)-vectors. Then using Lemma 2.2, we can prove the following:

THEOREM 3.2. *If the feasible regions of x_i are identical for each i, $i=1,2,\cdots,n$, then the optimal expected loss $R_n(x_n^*)$ is decreasing in n.*

Using Theorem 3.2, we can find the optimal experimental design x^* in a two dimensional search over the feasible region Ξ_n. We would choose n as large as possible and search for optimal $(\bar{x} - x_0, s_x)$ that minimizes

$$R_n(x) = E_{Y|x} E_{y_f|Y,x} \underset{d}{Min} E_{x_f|y_f,Y,x} [\ell(d, x_f)|y_f, Y, x].$$

In geneal, it is exceedingly difficult to calcult to calcuate R(x) even for one particular experimental design x. If $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ is $N_{2q}\left[\begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} D_a & 0 \\ 0 & D_b \end{bmatrix}\right]$, x_f is $N(x_0, \sigma_0^2)$, and $\ell(d, x_f) = (d - x_f)^2$, —which will be assumed hereafter, —then the expected loss is

$$R(x) = E_{Y|x} E_{y_f|Y,x} [Var(x_f|y_f, Y, x], \tag{3.6}$$

with optimal decision regarding x_f

$$d^*(y_f, Y, x) = E[x_f|y_f, Y, x].$$

Let $w_f = y_f - \mu_\alpha$. Then the posterior distribution of x_f, (3. 5), can be rewritten as

$$p(x_f|y_f, Y, x) \propto |2\pi(AS^{-1}A^T + \Gamma)|^{-\frac{1}{2}} \exp\{-1/2[(x_f - x_0)^2/\sigma_0^2 + [w_f - (x_f - x_0)\mu_\beta]^T (AS^{-1}A^T + \Gamma)^{-1} [w_f - (x_f - x_0)\mu_\beta]]\}.$$

Since the posterior variance-covariance matrix of $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, S^{-1}, does not depend on the test result Y, it follows that w_f and μ_β are sufficient for x_f with respect to (y_f, Y), i.e., x_f is independent of (y_f, Y) given (w_f, μ_β, x). Therefore, the posterior distribution of x_f is

$$p(x_f|y_f, Y, x) = p(x_f|w_f, Y, x),$$

and hence R(x) can be expressed in terms of w_f and μ_β as

$$R(x) = E_{\mu_\beta|x} E_{w_f|\mu_\beta, x} [Var(x_f)|w_f, \mu_\beta, x]. \tag{3.7}$$

It follows that

$$R(x) = E_{\mu_\beta|x}\, E_{w_f|\mu_\beta x}\, E_{x_f|w_f\mu_\beta x}\,(x_f^2|w_f,\ \mu_\beta, x) \ -\ E_{\mu_\beta|x}\, E_{w_f|\mu_\beta x}\,[E_{x_f|w_f\mu_\beta x}\,(x_f|w_f,\ \mu_\beta, x)]^2.$$

Since x_f is independent of w_f and μ_β, we can explicitly evaluate the first term so that

$$R(x) \ = \ x_0^2+\sigma_0^2-\int\int p(\mu_\beta|x)\ \frac{[\int x_f p(x_f)p(w_f|x_f,\ \mu_\beta,x)dx_f]^2}{[\int p(x_f)p(w_f|x_f,\ \mu_\beta,x)dx_f]}\ dw_f d\mu_\beta. \tag{3.8}$$

We can see from (3.8) that $(2q+1)$ nested integrals are necessary to calculate R(x) for one design x.

It can be shown that the distribution of μ_β given x and that of w_f given (x_f,μ_β,x) are normal, so that R(x) can be calculated by the densities of x_f, w_f and μ_β. This leads to the important result that $R(x) = R(n,\ \bar{x}-x_0,\ s_x)$ is symmetric in $\bar{x}-x_0$ for fixed n and s_x^2 (see Kim [12] and Barlow, Mensing and Smiriga [2]). This result further reduces our design space so that we only have to examine R(x) in the region of the intersection of $0\le\bar{x}-x_0\le s_x$ and any given constraints for fixed n.

3.3. Optimal Design Proposed

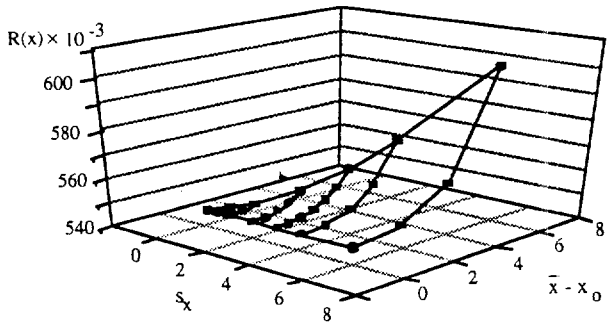
Suppose we are uncertain about both α and β. For, say, Design (I) such that $x_1=x_2=\cdots x_n=x_0$, the likelihood (3. 1) becomes

$$L(\alpha,\beta|data)\ \propto\ \exp\{-1/2\ tr\Gamma^{-1}[n(\alpha-a)(\alpha-a)^T-2(\alpha-a)1^TU]\}.$$

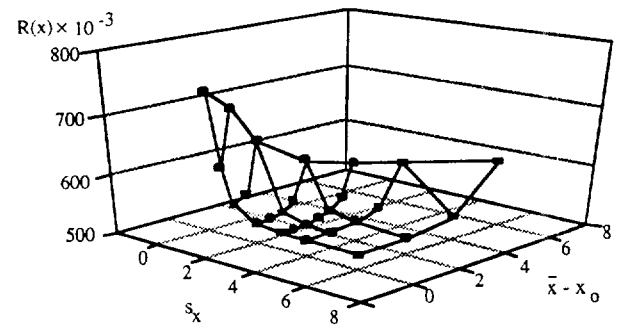
The likelihood does not involve β, which means that with Design (I), the calibration experiment provides no information about β. Intuitively, if β is unknown, Design (I) is a local maximum for R(x) since any design with x_i near x_0 will provide information about β and hence reduce the final expected loss.

Consider Design (II), $\bar{x}-x_0 = 0$, where x_0 is the prior mean of x_f. In this case, the calibration experiment reflects the prior belief on x_f, the future x-values. Clearly, Design (II) is better than the design with $\bar{x}\neq x_0$ because of the reflection of the prior for x_f. For this argument, only the prior mean needs to be specified. Technically, Design (I) is a special case of (II), and the reason why (I) is bad is that (I) has no variation in the design, i.e., $\sum_{i=1}^{n}(x_i-x_0)^2=0$.

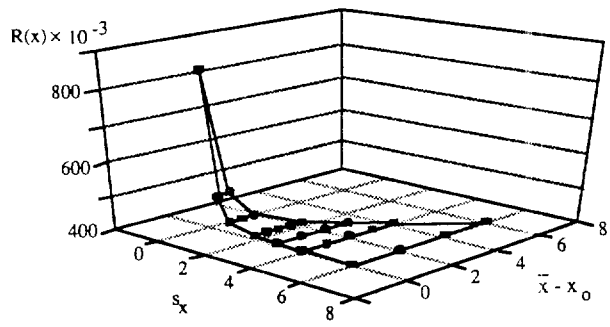
One-point design, Design (III), is the one that is either $\bar{x}-x_0 = s_x$ or $\bar{x}-x_0 = -s_x$. Elfving [9] and Chaloner [7] showed that in a linear regression framework, the one-point de-



(a) n=5 σ=1
x_0=0 σ_0=1
a=1 σ_a=2
b=1 σ_b=0.1



(a) n=5 σ=1
x_0=0 σ_0=1
a=1 σ_a=1
b=1 σ_b=1.2



(a) n=5 σ=1
x_0=0 σ_0=1
a=1 σ_a=0.1
b=1 σ_b=2

Figure 3.1 A three dimensional plot of R(x)

sign is optimal for estimating the linear combination of the coefficient β. Their linear models for which one-point designs are optimal correspond to ours with the case when α is known. When both α and β are unknown, however, the one-point design is no longer optimal as we can see from Figure 3. 1. Figure 3. 1 illustrates the intuition argued above, where the typical plots for the univariate case ($q=1$) using a numerical calculation are shown.

Table 3.1 Calculated values for Figure 3.1

Each cell showing R(x) and R'(x) ; $n=5$, $\sigma=1$, $x_0=0$, $\sigma_0=1$, a=1, b=1

(a) $\sigma_a=2$, $\sigma_b=0.1$, Lower bound=0.497512, EV(∞)=0.554507

	s_x					
$\bar{x}-x_0$	0	0.5	1	2	3	4
0	0.546989 0.543478	0.546925 0.543478	0.546919 0.543478	0.546904 0.543478	0.546873 0.543478	0.546849 0.543478
0.5		0.547393 0.543950	0.547370 0.543963	0.547293 0.543876	0.547199 0.543807	0.547110 0.543742
1			0.548788 0.545358	0.548519 0.545113	0.548205 0.544822	0.547914 0.544554
2				0.554207 0.550847	0.552745 0.549403	0.551454 0.548128
3					0.562813 0.559548	0.558813 0.555556
4						0.574002 0.570866

(b) $\sigma_a=1$, $\sigma_b=1.2$, Lower bound=0.290698, EV(∞)=0.528227

	s_x					
$\bar{x}-x_0$	0	0.5	1	2	3	4
0	0.735003 0.538462	0.615367 0.538462	0.557700 0.538462	0.534188 0.538462	0.529094 0.538462	0.527244 0.538462
0.5		0.707189 0.576087	0.568045 0.546275	0.535755 0.540336	0.529702 0.539288	0.527567 0.538925
1			0.656282 0.618457	0.541469 0.547242	0.531688 0.541996	0.528587 0.540386
2				0.620941 0.649899	0.543306 0.557984	0.533559 0.547522
3					0.614979 0.658631	0.547321 0.567355
4						0.613591 0.662019

(c) $\sigma_a=0.1$, $\sigma_b=2$, Lower bound=0.166667, EV($\propto$)=0.404065

	s_x					
$\bar{x}-x_0$	0	0.5	1	2	3	4
0	0.848897 0.502370	0.505219 0.502370	0.437027 0.502370	0.413937 0.502370	0.409227 0.502370	0.407536 0.502370
0.5		0.508839 0.502467	0.437832 0.502397	0.413966 0.502377	0.409233 0.502373	0.407538 0.502371
1			0.438494 0.502482	0.414052 0.502398	0.409254 0.502382	0.407547 0.502377
2				0.414412 0.502486	0.409337 0.502420	0.407578 0.502398
3					0.409482 0.502487	0.407632 0.502434
4						0.407711 0.502487

A simple observation on the posterior distribution of (α,β) in the univariate case, $q=1$, will give us an attractive insight. For the univariate case, σ^2 is the error variance and (σ_a^2, σ_b^2) are the prior variances of (α,β). The posterior (α,β) given the calibration experiments has a bivariate normal distribution with parameters $(\mu_\alpha, \mu_\beta, \sigma_\alpha^2, \sigma_\beta^2, \rho_{\alpha\beta})$. As in the multivariate case, the variances and covariance depend on the experimental design only. Consider the case $n \geq 2$ and fixed. It can be easily verified from the posterior distribution of (α,β) which is derived in [2], that for fixed $\bar{x}-x_0$, both σ_α^2 and σ_β^2 decrease as s_x increases, and for fixed s_x, they decrease as $|\bar{x}-x_0|$ decreases. Further, both σ_α^2 and σ_β^2 have their smallest values when the design is such that $|\bar{x}-x_0|=0$ and s_x as large as possible, which indicates that this design gives the most information on (α,β). Thus, we might say that this design, if feasible, is the most desirable one. The case $n=1$ seems to be a pathological case because $|\bar{x}-x_0|$ and s_x are identically equal to $|x_1-x_0|$ so that we cannot make one large and the other small at the same time. In fact for $n=1$, as $|x_1-x_0|$ increases, σ_α^2 increases and σ_β^2 decreases.

The optimal solution procedure can be proposed based on the preceding argument as following:

CONJECTURE . *For* $n>1$, $R(x) = R(n, |\bar{x}-x_0|, s_x)$ *is decreasing in* s_x *and increasing in* $|\bar{x}-x_0|$ *for others fixed.*

4. APPROXIMATIONS IN CONDITIONAL COVARIANCE MATRIX

In this section, we introduce an idea of approximating the conditional covariance matrix Q of $\begin{bmatrix} x_f \\ y_f \end{bmatrix}$ given (α,β) to derive the optimal design for the approximate problem. Given (α,β), the conditional distribution of y_f given x_f *is* $N_q(\alpha+(x_f-x_0)\beta,\ \Gamma)$. Therefore, the joint distribution of $\begin{bmatrix} x_f \\ y_f \end{bmatrix}$ given (α,β),

$$p(x_f y_f|\alpha,\beta)=p(y_f|x_f,\alpha,\beta)\ p(x_f),$$

is a $(q+1)$-variate normal distribution with mean vector $\begin{bmatrix} x_0 \\ \alpha \end{bmatrix}$ and covariance matrix

$$Q = \begin{bmatrix} \sigma_0^2 & \sigma_0^2\beta^T \\ \sigma_0^2\beta & \Gamma+\sigma_0^2\beta\beta^T \end{bmatrix}. \tag{4.1}$$

This joint distribution can be obtained by collecting terms and matching coefficients for normal densities.

The conditional covariance matrix Q involves the conditioning variable β, which is the crux of the difficulties in this problem. If we can replace β in Q by its prior mean b, we can find a near-optimal design under a mild condition. Let's approximate Q by

$$P = \begin{bmatrix} \sigma_0^2 & \sigma_0^2 b^T \\ \sigma_0^2 b & \Gamma+\sigma_0^2 bb^T \end{bmatrix}, \tag{4.2}$$

where we replace β by b. This approximation will be accurate if we are pretty sure that β is near b, that is, $D_b \approx 0$.

Since $\begin{bmatrix} x_f \\ y_f \end{bmatrix}$ given (α,β) is approximately

$$\begin{bmatrix} x_f \\ y_f \end{bmatrix} \Big| \ \alpha,\beta \ \sim \ N_{q+1}\left[\begin{bmatrix} x_0 \\ \alpha \end{bmatrix},\ P\right],$$

and

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \Big| \ Y,x \ \sim \ N_{2q}\left[\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix},\ S^{-1} = \begin{bmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{bmatrix}\right],$$

the predictive distribution of $\begin{bmatrix} x_f \\ y_f \end{bmatrix}$ can be derived as

$$\begin{bmatrix} x_f \\ y_f \end{bmatrix} \Big| \ Y,x \ \sim \ N_{q+1}\left[\begin{bmatrix} x_0 \\ \mu_\alpha \end{bmatrix},\ \begin{bmatrix} \sigma_0^2 & \sigma_0^2 b^T \\ \sigma_0^2 b & S^{11}+\Gamma+\sigma_0^2 bb^T \end{bmatrix}\right],$$

where $S^{11} = V(\alpha|Y,x)$ which involves the design x. Then, x_f given (y_f,Y,x) is normal and its variance is

$$Var(x_f|y_f,Y,x) = \sigma_0^2 - \sigma_0^4 b^T(S^{11}+\Gamma+\sigma_0^2 bb^T)^{-1}b.$$

Since this does not depend on y_f or Y, it follows that

$$R(x) = E_{Y|x}E_{y_f|Y,x}[Var(x_f|y_0,Y,x)]$$
$$\approx \sigma_0^2-\sigma_0^4 b^T[S^{11}(x)+\Gamma+\sigma_0^2 bb^T]^{-1}b = R'(x). \quad (4.3)$$

Minimizing (4.3) is equivalent to maximizing $b^T M(x)^{-1}b$ where

$$M(x) = S^{11}(x)+\Gamma+\sigma_0^2 bb^T \quad (4.4)$$

If $b^T b > 0$ we can characterize the optimal design for the approximate problem. We will use the notation $A > 0$ (≥ 0) to mean that A is positive (semi) definite.

THEOREM 4.1. *IF $D_b \approx 0$, then with the approximation (4.2), the optimal calibration design for fixed $n \geq 2$ is to choose x^* such that*

$$S^{11}(x)-S^{11}(x^*) \geq 0,$$

for all $x \in \Xi_n$. Furthermore, if $S^{11}(x)>0$ for all x, then a near optimal design is the one that is $\bar{x}-x_0 = 0$ and $s_x \neq 0$.

PROOF. Fix $n \geq 2$. Let $x \in \Xi_n$ and x^* be such that $S^{11}(x)-S^{11}(x^*) \geq 0$. Since $b^T b \geq 0$, $\sigma_0^2 bb^T \geq 0$. Let us denote $\lambda_k(A)$ the k-th largest eigenvalue of matrix A. Since $\Gamma > 0$, and $S^{11}(x)$ and $\sigma_0^2 bb^T$ are ≥ 0, it follows from Beckenbach and Bellman [3][p73] that

$$\lambda_k[M(x)] \geq \lambda_k[S^{11}(x)+\Gamma] \geq \lambda_k(\Gamma) > 0, \quad k=1,2,3,\cdots,q,$$

which implies $M(x) > 0$.

Since $S^{11}(x)-S^{11}(x^*) \geq 0$, we have $M(x)-M(x^*) \geq 0$. It can be shown that for $M(x)>0$ and $M(x^*)>0$, if $M(x)-M(x^*) \geq 0$ then $M(x^*)^{-1}-M(x)^{-1} \geq 0$. [3]. Therefore,

$$b^T[M(x^*)^{-1}-M(x)^{-1}]b \geq 0,$$

or

$$b^T M(x^*)^{-1}b \geq b^T M(x)^{-1}b,$$

hence by (4.3) and (4.4), x^* is optimal.

Now suppose $S^{11}(x) > 0$. And let $\xi \in \Xi_n$ be such that $\bar{\xi}-x_0=0$ *and* $s_\xi \neq 0$, where $\bar{\xi} = \sum \xi_i/n$. Note that, from (3.3),

$$S^{11}(x) = (S_{11}-S_{12}S_{22}^{-1}S_{21})^{-1}$$
$$= [D_b^{-1}+n\Gamma^{-1}-n^2(\bar{x}-x_0)^2\Gamma^{-1}S_{22}^{-1}\Gamma^{-1}]^{-1}$$

Since $D_b^{-1} \geq 0$ and $\Gamma^{-1} > 0$, $S_{22} = D_b^{-1}+ns_x^2\Gamma^{-1}$ is positive definite, and so is S_{22}^{-1}, hence $\Gamma^{-1}S_{22}^{-1}\Gamma^{-1}$ is positive definite. Thus, we have

$$n^2(\bar{x}-x_0)^2\ \Gamma^{-1}\ S_{22}^{-1}\ \Gamma^{-1}\geq 0,$$

or $$[D_a^{-1}+n\Gamma^{-1}]-[D_a^{-1}+n\Gamma^{-1}-n^2(\bar{x}-x_0)^2\ \Gamma^{-1}\ S_{22}^{-1}\ \Gamma^{-1}]\geq 0,$$

or $$[S^{11}(\xi)]^{-1}-[S^{11}(x)]^{-1}\geq 0.$$

Since both terms are positive definite, it follows that

$$S^{11}(x)-S^{11}(\xi)\geq 0\ ,$$

hence ξ is optimal. □

For the univariate case, (4.3) reduces to

$$\begin{aligned} R^1(x) &= \sigma_0^2-\sigma_0^4 b^2/(\sigma_x^2+\sigma^2+\sigma_\beta^2 b^2) \\ &= \left[\frac{1}{\sigma_0^2}+\frac{b^2}{\sigma_0^2+\sigma_x^2}\right]^{-1}. \end{aligned} \tag{4.3a}$$

$R^1(x)$ depends on x only through σ_x^2, so that the near optimal design is to minimize σ_x^2, which is the design with $\bar{x}-x_0=0$. This agrees with the result of Theorem 4.1. Since the approximation $R^1(x)$ is good for a small σ_b, it is worthwhile to compare $R^1(x)$ with $R(x)$ for $\sigma_b=0$, whose formula is given in Kim[1988]. The two formulas are the same, but the values of σ_x^2 being applied are different. For $R^1(x)$, σ_x^2 depends on the prior variance of β, while for $R(x)$ when $\sigma_b=0$, it does not. In fact, when $\sigma_b=0$, R(x) depend on n only. Although $R^1(x)$ does not depend on the posterior variance of β, the significance of the approximation $R^1(x)$ is the implication that when σ_b is relatively small, the value of $\bar{x}-x_0$ is more important than s_x in determining the optimal design, which agrees with the results of the numerical calculation in Figure 3.1. (a).

5. NUMERICAL EXAMPLES

In Figure 3. 1, we showed three plots of R(x) from numerical calculations. Each plot represents different situation of prior variance of α and β. In this section, we investigate the three cases in detail and compare R(x) with the approximation $R^1(x)$ given in (4.3.a). Because of the complexity in the numerical integration, we will consider the univariate case.

To show the effects of the different prior variances of α and β clearly, we use the same values for some parameters. For all three cases, we have fixed the error variance, the prior mean and variance of x_f and the prior means of α and β, as shown in the figure. The number

of grid points in one numerical integration is 80. The selection of $n=5$ is arbitrary, but when n is large, R(x) is small and has little variation. The choice of σ and σ_x are not major factors for the structure of R(x). The three cases are (a) $\sigma_a = 2$, $\sigma_b = 0.1$; (b) $\sigma_a = 1$, $\sigma_b = 1.2$; and (c) $\sigma_a = 0.1$, $\sigma_b = 2$.

In Figure 3.1, (a) shows the three dimensional plot of R(x) when σ_b is much less than σ_a, for which we have little uncertainty in β. This is almost the case of $\sigma_b=0$, where R(x) is shown to depend on n only. Here, R(x) has some variation in $|\bar{x}-x_0|$ but negligible one in s_x. The case (c) shows the plot when σ_a is much less than σ_b. This is almost the case of $\sigma_b=0$, where R(x) is decreasing in s_x. Indeed, R(x) here is almost a constant in $|\bar{x}-x_0|$ for a fixed s_x, and decreasing in s_x. The case (b) is in between these. Notice in all three cases, that for fixed s_x, R(x) gets smaller values as $|\bar{x}-x_0|$ decreases, and for fixed $|\bar{x}-x_0|$, R(x) gets smaller as s_x increases, which coincides with the behavior of σ_z^2 and σ_β^2. Furthermore, R(x) is the smallest when $|\bar{x}-x_0| = 0$ and s_x as large as possible (here $s_x=4$) in all three cases.

All three cases agree with the conjecture given in Section 3.3 that an optimal design is such that $|\bar{x}-x_0| = 0$ and s_x largest possible. For each design in Table 3.1, there are two numbers corresponding to R(x) and $R^1(x)$. We also have, as a reference, the value EV(∞) of R(x) for the design that any level of the calibration experiment becomes infinity. This design corresponds $|\bar{x}-x_0|\to\infty$ and $s_x\to\infty$, so that this design is never optimal. We also have, in Table 3.1, the lower bound (2. 7) of Section 2 to see whether the approximations are acceptible. No value of the approximation violates this bound. All three cases seem to be consistent with the results for the special cases, except in (a) for $\sigma_b\approx 0$, R(x) depends on $|\bar{x}-x_0|$ more than we expected. The approximation $R^1(x)$ is more accurate for $\sigma_b\approx 0$, as we expected. Optimal designs using $R^1(x)$ are consistent with the result obtained in the previous section for all three cases, namely the ones with $|\bar{x}-x_0| = 0$.

REFERENCES

[1] Barlow, R. E., R. W. Mensing and N. G. Smiriga, "Using Influence Diagrams to solve a Calibration Problem," *Proc. International Symp. on Probability and Bayesian Statistics*, Innsbruck, Austria, September 23-26, (1986).

[2] Barlow, R. E., R. W. Mensing and N. G. Smiriga, "Computing the Optimal Design for a Calibration Experiment," *Journal of Statistical Planning and Inference*, Vol. 29(1991), pp5-19.

[3] Beckenbach, E. F. and R. Bellman, *Inequalities*, *Springer-Verlag*, New York, 1965.

[4] Brooks, R. J., "Optimal Regression Design for Control in Linear Regression," *Biometrika*, Vol. 64 (1977), pp319-325.

[5] Brown, P. J., "Multivariate Calibration," (with discussion), *Journal of Royal Statistical Society*, Ser B, Vol. 44 (1982), pp287-321.

[6] Buonaccorsi, J. P., "Design Considerations for Calibration," *Technometrics*, Vol. 28 (1986), pp149-155.

[7] Chaloner, K., "Optimal Bayesian Experimental Design for Linear Models," *Annals of Statistics*, Vol. 12 (1984), pp283-300.

[8] DeGroot, M. H., *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.

[9] Elfving, G., "Optimum Allocation in Linear Regression Theory," *Annals of Mathematical Statistics*, Vol. 23 (1952), pp255-262.

[10] Hoadley, B., "A Bayesian Look at Inverse Linear Regression," *Journal of American Statistical Association*, Vol. 65 (1970), pp356-369.

[11] Hunter, W. G. and W. F. Lamboy, "A Bayesian Analysis of the Linear Calibration Problem," (with discussion), *Technometrics*, Vol 23 (1981), pp323-350.

[12] Kim, S. C. *Bayesian Calibration Experimental Designs Based on Linear Models*, Ph. D. Thesis, Dept. of Industrial Engineering and Operations Research, University of California, Berkeley, 1988.

[13] Raiffa, H. and R. Schlaifer, *Applied Statistical Decision Theory*, MIT Press, Cambridge, Mass. 1961.