This research is being partially supported by Department of Health, Education, and Welfare Grant No. 5T01 ES00133, and by National Institutes of Health, Division of General Medical Sciences, Grant No. 5T01 GM00038.

A GENERALIZATION OF THE BOX-DRAPER APPROACH TO MEAN SQUARE ERROR ESTIMATION OF RESPONSE SURFACES: PRELIMINARY REPORT

by

Lawrence L. Kupper and Edward F. Meydrech

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 796

February 1972

ABSTRACT

Box and Draper [2,3] considered the problem of choosing a response surface design to minimize integrated mean square error J when the true response $\eta = \underset{\sim}{x}_1'\underset{\sim}{\beta}_1 + \underset{\sim}{x}_2'\underset{\sim}{\beta}_2$, represented as a polynomial of degree d_2 in p variables, is approximated by a polynomial $\hat{y} = \underset{\sim}{x}_1'\underset{\sim}{b}_1$ of lower degree d_1 for which the coefficient vector $\underset{\sim}{b}_1$ is the vector $\hat{\underset{\sim}{\beta}}_1$ of usual least squares estimates. Based on motivation supplied by Blight [1], Korts [9] and others, the use of a generalized estimator of the form $\underset{\sim}{b}_1 = \underset{\approx}{K}\hat{\underset{\sim}{\beta}}_1$, where $\underset{\approx}{K}$ is a diagonal matrix of appropriately chosen constants, is proposed. The choice of $\underset{\approx}{K}$ which minimizes J will depend on the unknown elements of the coefficient vectors $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$. This problem can be avoided by considering a modified criterion J^*, the expected value of J with respect to a prior distribution of these unknown parameters. For the case $d_1=1$, $d_2=2$, p=1, it is shown that the $\underset{\approx}{K}$ which minimizes J^* depends on the choice of prior only through its second moments and that a smaller value of J^* than when $\underset{\approx}{K}=\underset{\approx}{I}$ (the Box-Draper case) is attained for <u>any</u> choice of design. It is seen that the use of this generalized estimator offers the same flexibility as the Karson-Manson-Hader approach [8] in the sense that J^* can be even further reduced in value by an appropriate choice of experimental design.

1. INTRODUCTION

Response surface analysis is concerned with a functional relationship

$$\eta = f(\xi_1,\xi_2,\ldots,\xi_p;\ \theta_1,\theta_2,\ldots,\theta_r) = f(\underset{\sim}{\xi};\ \underset{\sim}{\theta})$$

between a response η and p non-stochastic factors $\xi_1,\xi_2,\ldots,\xi_p$; geometrically, $f(\underset{\sim}{\xi};\ \underset{\sim}{\theta})$ represents a surface in a (p+1)-dimensional space whose coordinates are the p ξ's and η. The shape of this surface can be assumed to depend on parameters $\theta_1,\theta_2,\ldots,\theta_r$, the values of which are usually unknown and must be estimated from experimental data typically consisting of N measured values of η corresponding to N specified combinations of levels of the factors

$$\underset{\sim}{\xi}_u' = (\xi_{1u},\xi_{2u},\ldots,\xi_{pu}),\quad u=1,2,\ldots,N.$$

The set of points $\{\underset{\sim}{\xi}_u\}_{u=1}^{N}$ is called the experimental design. Since the nature of the variables whose levels are represented by the $\{\xi_{iu}\}$ will change from one application to another, it is useful to define the experimental design in terms of "standardized" variables $\{x_{iu}\}$ of the general form

$$x_{iu} = (\xi_{iu}-\bar{\xi}_i)/S_i,\quad i=1,2,\ldots,p,\ u=1,2,\ldots,N,$$

where $\bar{\xi}_i = \frac{1}{N}\sum_{u=1}^{N}\xi_{iu}$ and S_i is a scale factor. Note that $\sum_{u=1}^{N} x_{iu} = 0$, i=1,2,...,p, so that the center of interest of the ξ's is the origin of the x's.

In practice, the form of the function $f(\underset{\sim}{\xi};\ \underset{\sim}{\theta})$ is generally unknown and the recommended procedure (see Box and Wilson [4]) is to approximate $f(\underset{\sim}{\xi};\ \underset{\sim}{\theta})$ by a low order polynomial. The coefficients of this polynomial are then the unknown parameters and can be estimated by standard least squares methods. The properties of such a polynomial graduating function are influenced by the choice of experimental design and, initially, criteria for judging the goodness of designs were

largely concerned with variance considerations. The question of bias due to inadequacy of the approximating polynomial was given only minor attention. In reality, of course, a graduating function such as a polynomial will always fail, to some extent, to represent the true relationship exactly.

More recently, in an effort to take "bias error" as well as "variance error" into account, Box and Draper (BD) [2,3] and Karson, Manson, and Hader (KMH) [8] adopted mean square error integrated over some region of interest $\mathcal{X}$ in the standardized space as a basic design criterion. In particular, if $\hat{y}(\underset{\sim}{x})$ denotes the value of the estimated response at the point $\underset{\sim}{x}' = (x_1, x_2, \ldots, x_p)$ in $\mathcal{X}$ where this value is obtained by fitting (not necessarily by least squares) a polynomial of degree $d_1 (\geq 1)$ in $x_1, x_2, \ldots, x_p$, and if $\eta(\underset{\sim}{x})$ is the true response at $\underset{\sim}{x}$ given exactly by a polynomial of higher degree d_2, then the above authors concerned themselves with the problem of determining the $\binom{p+d_1}{p}$ coefficients of $\hat{y}(\underset{\sim}{x})$ so as to minimize the multiple integral

$$J = \frac{N\Omega}{\sigma^2} \int_{\mathcal{X}} E[\hat{y}(\underset{\sim}{x}) - \eta(\underset{\sim}{x})]^2 d\underset{\sim}{x}, \tag{1}$$

where $\Omega^{-1} = \int_{\mathcal{X}} d\underset{\sim}{x}$ and $d\underset{\sim}{x} = dx_1 dx_2 \ldots dx_p$. The quantity J is the mean square error at $\underset{\sim}{x}$ averaged uniformly over $\mathcal{X}$ and normalized with respect to the number of observations and the experimental error variance σ^2. To illustrate the partition of J into a variance component and a bias component, (1) can be written as

$$J = V + B,$$

where

$$V = \frac{N\Omega}{\sigma^2} \int_{\mathcal{X}} \mathrm{Var}[\hat{y}(\underset{\sim}{x})] d\underset{\sim}{x} \tag{2}$$

and

$$B = \frac{N\Omega}{\sigma^2} \int_X [E\hat{y}(\underset{\sim}{x}) - \eta(\underset{\sim}{x})]^2 d\underset{\sim}{x}. \tag{3}$$

As expected, the minimization of J by choice of design depends on the relative magnitudes of the V and B contributions.

Using standard least squares estimates for the coefficients in the fitted polynomial and taking X to be some simple region such as a hypersphere [2,3] or hypercube [6], Box and Draper and Draper and Lawrence were led to the somewhat unexpected conclusion that unless the V contribution was many times larger than the B contribution, the optimal experimental designs for minimizing J were remarkably close to those obtained by ignoring V completely and minimizing B alone. Motivated by these results, Karson, Manson, and Hader developed an alternative approach which entailed using a method of estimation aimed directly at minimizing B and then employing any additional flexibility in the choice of design to minimize V. With this procedure, they were able to find designs with smaller J than those suggested by Box and Draper. One drawback to their "minimum bias estimator" is that, in most cases, its use requires that one obtain least squares estimates of <u>all</u> the $\binom{p+d}{p}$ coefficients in $\eta(\underset{\sim}{x})$. In this sense, their approach is somewhat less than practical.

2. ALTERNATIVE APPROACH

Blight [1] considered the problem of mean square error (MSE) estimation using an estimator which was a constant multiple of an unbiased estimator of the parameter of interest, and he gave sufficient conditions for uniform improvement (in terms of MSE) over unbiased estimation. His general approach was motivated by the fact that such estimators have been shown to be uniformly better than the ordinary unbiased estimator of powers of the standard deviation in normal samples (e.g., see Markowitz [12], Stein [13], Stuart [14]). For the

linear model $\underset{\sim}{y} = \underset{\sim}{x}\beta + \underset{\sim}{e}$ with the standard assumptions, Kupper [10] has noted that the linear estimator $b = \underset{\sim}{c}'\underset{\sim}{y}$ which minimizes $MSE(b) = \underset{\sim}{c}'\underset{\sim}{c}\sigma^2 + (1-\underset{\sim}{c}'\underset{\sim}{x})^2\beta^2$ is of the form $k\hat{\beta}$, where $k = \beta^2/[\beta^2 + (\underset{\sim}{x}'\underset{\sim}{x})^{-1}\sigma^2]$ and $\hat{\beta} = (\underset{\sim}{x}'\underset{\sim}{x})^{-1}\underset{\sim}{x}'\underset{\sim}{y}$ is the best linear unbiased estimator of β. In this same setting, Korts [9] demonstrated that the b which maximizes the minimum over $\{\beta: |\beta| \leq M\}$ of $Pr(\beta-\varepsilon \leq b \leq \beta+\varepsilon)$ is also a multiple of $\hat{\beta}$, this multiple being $2[1 + (1+2\delta)^{\frac{1}{2}}]^{-1}$, where $\delta = \frac{(\underset{\sim}{x}'\underset{\sim}{x})^{-1}\sigma^2}{M\varepsilon} \log_e \left(\frac{M+\varepsilon}{M-\varepsilon}\right)$, $0<\varepsilon<M$. In the multivariate case $\underset{\sim}{y} = \underset{\sim}{\underset{\sim}{X}}\underset{\sim}{\beta} + \underset{\sim}{e}$, Korts obtained (under a mild restriction) the analogous result that the elements of $\underset{\sim}{b} = \underset{\sim}{C}'\underset{\sim}{y}$ which jointly satisfy his "closeness" criterion are the elements of $\underset{\sim}{K}\hat{\underset{\sim}{\beta}}$, where $\underset{\sim}{K}$ is a specified diagonal matrix and $\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y}$.

We can relate the implications of the above work to the problem of minimizing integrated mean square error in the following way. In the setting of Section 1, let $\eta(\underset{\sim}{x}) = \underset{\sim}{x}_1'\underset{\sim}{\beta}_1 + \underset{\sim}{x}_2'\underset{\sim}{\beta}_2$ and let $\hat{y}(\underset{\sim}{x})$ be of the form $\underset{\sim}{x}_1'\underset{\sim}{b}_1$. The vector $\underset{\sim}{x}_1$ is made up of terms required for the polynomial of degree d_1; the vector $\underset{\sim}{x}_2$ consists of the additional higher order terms required to represent the polynomial of degree d_2, while $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$ are the corresponding vectors of unknown coefficients. With this framework in mind, we define a new estimator

$$\underset{\sim}{b}_1 = \underset{\sim}{K}(\underset{\sim}{X}_1'\underset{\sim}{X}_1)^{-1}\underset{\sim}{X}_1'\underset{\sim}{y},$$

where $\underset{\sim}{K}$ is a size $\binom{p+d_1}{p}$ diagonal matrix of appropriately chosen constants, $\underset{\sim}{X}_1$ is the matrix of values taken by the terms in $\underset{\sim}{x}_1$ for the N experimental combinations of the independent variables, and $\underset{\sim}{y}$ is the column vector of the measured values $y_u = \eta(\underset{\sim}{x}_u) + e_u$, $u=1,2,\ldots,N$. We make the usual assumptions that $E(e_u)=0$, $E(e_u^2)=\sigma^2$, $E(e_u e_{u'})=0$ for $u \neq u'$. Note that in the special case when $\underset{\sim}{K}=\underset{\sim}{I}$, our generalized estimator $\underset{\sim}{b}_1$ is simply the least squares estimator employed by Box and Draper.

As would be expected, the particular choice of $\underset{\sim}{K}$ (i.e., the choice of the diagonal elements of $\underset{\sim}{K}$) which minimizes (1) will depend, unfortunately, on the unknown elements of the parameter vectors $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$ and also on σ^2. To get around this obstacle, one possible alternative would be to consider minimizing the maximum of J over a restricted parameter space (a la Korts' maximin procedure). Another approach (somewhat Bayesian in nature) would involve minimizing by choice of $\underset{\sim}{K}$ the quantity obtained by averaging J over a prior distribution of the unknown parameters. This second approach, which is the one we shall consider in this preliminary report, is not only much more amenable to analysis, but it also leads (at least in the examples we have studied) to meaningful results whose interpretations do not necessitate an exact specification of the prior distribution. This concept of taking the expected value of J with respect to some prior distribution is somewhat similar in motivation to the notion that Box and Draper [2] had of averaging J over all orthogonal rotations of the response surface.

If $g(\underset{\sim}{\beta})$ denotes a prior distribution for the elements of $\underset{\sim}{\beta}' = (\underset{\sim}{\beta}_1', \underset{\sim}{\beta}_2')$ with domain $\mathcal{B}$, then the general criterion of interest to us is

$$J^* = E(J) = \int_{\mathcal{B}} J g(\underset{\sim}{\beta}) d\underset{\sim}{\beta}.$$

In our particular situation, the primary objective is to choose $\underset{\sim}{K}$ to minimize

$$J^* = V^* + B^*$$

when

$$V^* = V = N\Omega \int_{\mathcal{X}} \underset{\sim}{x}_1' \underset{\sim}{K} (\underset{\sim}{X}_1' \underset{\sim}{X}_1)^{-1} \underset{\sim}{K} \underset{\sim}{x}_1 \, d\underset{\sim}{x} \tag{4}$$

and

$$B^* = \frac{N\Omega}{\sigma^2} \int_{\mathcal{B}} \int_{X} [\underset{\sim}{x}_1' \underset{\sim}{K}(\underset{\sim}{\beta}_1 + \underset{\sim}{A}\underset{\sim}{\beta}_2) - (\underset{\sim}{x}_1'\underset{\sim}{\beta}_1 + \underset{\sim}{x}_2'\underset{\sim}{\beta}_2)]^2 d\underset{\sim}{x}\, g(\underset{\sim}{\beta}) d\underset{\sim}{\beta}, \tag{5}$$

where $\underset{\sim}{A} = (\underset{\sim}{X}_1'\underset{\sim}{X}_1)^{-1}\underset{\sim}{X}_1'\underset{\sim}{X}_2$ and $\underset{\sim}{X}_2$ is the counterpart of $\underset{\sim}{X}_1$ when $\underset{\sim}{x}_2$ replaces $\underset{\sim}{x}_1$. Note that V is neither a function of the elements of $\underset{\sim}{\beta}$ nor of σ^2.

Because of the complexity of the expressions (4) and (5), an explicit solution for $\underset{\sim}{K}$ for general d_1, d_2, and p is hard to come by. However, this is no real problem since the specific cases of most importance can be dealt with quite successfully. For example, the case $d_1=1$, $d_2=2$, $p=1$, which has received considerable attention in the literature [2,5,8], is considered in detail in the next section; investigation of certain multivariate ($p>1$) problems of interest to response surface workers is currently under way and the results will be reported in a later paper.

One important advantage of our procedure over that of Box and Draper is that it enjoys the same flexibility as the KMH approach with regard to the choice of experimental design. In particular, the value of J* for the optimal choice of $\underset{\sim}{K}$ can be even further reduced by selecting an appropriate design. This phenomenon will be illustrated in the following example where, interestingly enough, it is shown that our approach gives a smaller value of J* than when $\underset{\sim}{K}=\underset{\sim}{I}$ (the BD case) for <u>any</u> choice of design. The finding of such "uniform" improvement is not unexpected in view of the work of Blight and others cited earlier.

3. APPLICATION: FITTING A STRAIGHT LINE

We will illustrate the theory for the special case $d_1=1$, $d_2=2$, $p=1$.

For simplicity and without loss of generality, we take the region of interest X to be the closed interval [-1, 1]. We assume that the true response is given by the quadratic expression

$$\eta(x) = \beta_0 + \beta_1 x + \beta_2 x^2.$$

Our estimator of $\eta(x)$ takes the form

$$\hat{y}(x) = k_0\hat{\beta}_0 + k_1\hat{\beta}_1 x,$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the usual least squares estimators of intercept and slope calculated via N pairs of data points (x_u,y_u), u=1,2,...,N, for which $\sum_{u=1}^{N} x_u = 0$.

Let $\mu_j = \frac{1}{N}\sum_{u=1}^{N} x_u^j$ denote the j-th moment of the design points, j=2,3. Then, since $\mathrm{Var}[\hat{y}(x)] = k_0^2\,\mathrm{Var}(\hat{\beta}_0) + k_1^2x^2\,\mathrm{Var}(\hat{\beta}_1)$, it is easy to see from (2) that

$$V = k_0^2 + k_1^2/3\mu_2. \tag{6}$$

Similarly, if we note that $E(\hat{\beta}_0) = \beta_0 + \mu_2\beta_2$ and $E(\hat{\beta}_1) = \beta_1 + \mu_3\beta_2/\mu_2$, then some simple algebra enables us to put (3) in the form

$$\begin{aligned} B = \frac{4}{45}\alpha_2^2 &+ [k_0(\alpha_0+\mu_2\alpha_2) - (\alpha_0+\alpha_2/3)]^2 \\ &+ \tfrac{1}{3}[k_1(\alpha_1+\mu_3\alpha_2/\mu_2) - \alpha_1]^2, \end{aligned} \tag{7}$$

where $\alpha_i = \beta_i/(\sigma/\sqrt{N})$, i=0,1,2. The sampling error in $\hat{y}(x)$ is proportional to $\sigma/\sqrt{N}$ so that α_1^2, for example, is a "standardized" measure of the strength of the linear component of $\eta(x)$. When $k_0=k_1=1$, expressions (6) and (7) correspond exactly to expressions (17) and (23) on pages 627-8 of Box and Draper [2].

The sets of values of k_0 and k_1 which minimize (6) and (7) can be determined by inspection, and it is a straightforward exercise to show that J, the sum of (6) and (7), is minimized when

$$k_0 = \frac{(\alpha_0+\mu_2\alpha_2)(\alpha_0+\alpha_2/3)}{1+(\alpha_0+\mu_2\alpha_2)^2} \tag{8}$$

and

$$k_1 = \frac{\alpha_1\mu_2(\alpha_1+\mu_3\alpha_2/\mu_2)}{1+\mu_2(\alpha_1+\mu_3\alpha_2/\mu_2)^2} \tag{9}$$

It is worth mentioning that the value of J for k_0 and k_1 given by (8) and (9) is less than the corresponding value when $k_0=k_1=1$ (the BD case) for <u>any</u> choice of design for which $\mu_3=0$ (the optimal third moment value when $k_0=k_1=1$).

For a situation in which the $\{\alpha_i\}$ are known, the above results will be of considerable importance. In practice, however, the $\{\alpha_i\}$ are unknown. In such circumstances, Box and Draper were led to their "rotational average" concept mentioned earlier, and we are led to our J* criterion.

In order to implement our J* criterion, we need to specify a prior distribution for the unknown parameters β_0, β_1, and β_2. Let $g_i(\beta_i)$ represent a prior distribution for β_i with domain $\mathcal{B}_i$, i=0,1,2. It is standard practice to assume that

$$g(\beta_0,\beta_1,\beta_2) = \prod_{i=0}^{2} g_i(\beta_i).$$

Also, for most of the priors of interest to us, there is no loss in generality if we take $E(\beta_i) = 0$ for every i, in which case $\gamma_i = E(\alpha_i^2)$ is then the average over the prior on β_i of the standardized measure of the magnitude of β_i mentioned earlier. As an important example, if sufficient knowledge of the parameter space is available so that a constant M_i can be obtained for which it is known that $|\beta_i| \leq M_i$, then an obvious choice for a prior on β_i is the uniform distribution over $[-M_i, M_i]$. The use of such a "vague" or "non-informative" prior provides a formal way of expressing ignorance with regard to the value of the parameter over the range permitted. In other words, nothing is said about the value of the parameter, except the basic fact that it is restricted, by its

very nature, to lie within certain definite limits. An excellent discussion of the philosophy involved in specifying prior distributions for parameters can be found in the third chapter of Jeffreys [7].

Our problem now becomes one of choosing k_0 and k_1 to minimize $J^* = V+B^*$, where $V = k_0^2 + k_1^2/3\mu_2$, and, for B given by (7),

$$B^* = \int_{\mathcal{B}_0} \int_{\mathcal{B}_1} \int_{\mathcal{B}_2} B\, g_0(\beta_0) g_1(\beta_1) g_2(\beta_2) d\beta_0 d\beta_1 d\beta_2$$

$$= (k_0-1)^2\gamma_0 + \frac{1}{3}(k_1-1)^2\gamma_1 \qquad (10)$$

$$+ [(k_0\mu_2 - \tfrac{1}{3})^2 + k_1^2\mu_3^2/3\mu_2^2 + \tfrac{4}{45}]\gamma_2.$$

When there is no experimental error and all discrepancy arises solely from bias, then it is of interest to find the values of k_0 and k_1 which minimize (10); these values are

$$k_0 = (\gamma_0+\mu_2\gamma_2/3)/(\gamma_0+\mu_2^2\gamma_2)$$

and

$$k_1 = \gamma_1/(\gamma_1 + \mu_3^2\gamma_2/\mu_2^2).$$

To lower the value of (10) further by choice of design, it is clear that we should set $\mu_3=0$. It is then fairly easy to show that the optimal value for the second moment is $\mu_2 = 1/3$, in which case k_0 and k_1 above are both equal to 1. Thus, our methodology leads us back to the conclusions of Box and Draper [2] concerning the minimization of integrated squared bias.

With a little work, it can be proved that the minimization of J* requires that $\mu_3=0$ and that

$$k_0 = (\gamma_0 + \mu_2\gamma_2/3)/(1 + \gamma_0 + \mu_2^2\gamma_2) \tag{11}$$

and

$$k_1 = \mu_2\gamma_1/(1 + \mu_2\gamma_1). \tag{12}$$

For <u>any</u> choice of $\mu_2 \geq 1/3$, the value of J* for the above values of k_0 and k_1 can be shown to be less than the value of J* obtained via the BD approach ($k_0=k_1=1$); this is the "uniform" improvement property alluded to earlier. Also, note that (11) and (12) depend on the choice of the prior distribution only through its second moments.

The <u>optimal</u> choice for μ_2 can be found as follows. The sum of (6) and (10) can be written as

$$J^* - (\gamma_0+\gamma_1/3+\gamma_2/5) = -f_0(\mu_2)-f_1(\mu_2)$$

$$= -(\gamma_0+\mu_2\gamma_2/3)k_0 - \gamma_1 k_1/3,$$

where k_0 and k_1 are given by (11) and (12). Thus, we can equivalently find the value of μ_2 which maximizes $f(\mu_2) = f_0(\mu_2) + f_1(\mu_2)$. Now, it is clear that $f_1(\mu_2)$ is maximized when $\mu_2=1$, and it is easy to show that $f_0(\mu_2)$ has only one extremum point, a simple maximum occurring at $\mu_2 = \frac{1}{3}(1+1/\gamma_0)$. Thus, if $\gamma_0 \leq 1/2$, so that sampling error is more than twice the size of the variation in β_0, we should take $\mu_2=1$. In contrast to this result, typical values of μ_2 obtained by Box and Draper were closer to 1/3 (their "all bias error, no variance error" value) than to 1 (their "all variance error, no bias error" value), and such values of μ_2 were determined solely by the size of $N\beta_2^2/\sigma^2$, a

"strength of bias" measure. One possible reason for such differing results is that in our approach we are willing to tolerate a "structural" change (via k_0 and k_1) in the bias term in order to achieve a reduction in the variance contribution.

When $\gamma_0 > 1/2$, the optimal value of μ_2 must be obtained by solving $f'(\mu_2)=0$, which results in the fourth degree polynomial equation

$$\sum_{j=0}^{4} a_j \mu_2^j = 0,$$

where

$$a_4 = 6\gamma_1^2\gamma_2^2(\gamma_0 - 1/2),$$

$$a_3 = 2\gamma_1\gamma_2(9\gamma_0^2\gamma_1 + 6\gamma_0\gamma_2 - \gamma_1\gamma_2 - \gamma_0\gamma_1\gamma_2),$$

$$a_2 = 2a_3/\gamma_1 - 6\gamma_2[\gamma_1^2(1+\gamma_0)^2 + 3\gamma_0\gamma_2],$$

$$a_1 = a_2/2\gamma_1 - 6\gamma_1\gamma_2(\gamma_0-1/2) - 3\gamma_0\gamma_2^2/\gamma_1 - 9\gamma_0^2\gamma_1\gamma_2,$$

$$a_0 = -3(1+\gamma_0)[\gamma_1^2(1+\gamma_0) + 2\gamma_0\gamma_2].$$

From the structure of these coefficients, it follows from Descartes' Rule of Signs (see [11], p. 60) that for $\gamma_0 > 1/2$ the above polynomial has at most one positive root. Since $f'(\mu_2) < 0$ for $0 \leq \mu_2 \leq \frac{1}{3}(1+1/\gamma_0)$ and since $f'(\mu_2) > 0$ for μ_2 large enough, $f'(\mu_2)=0$ has <u>exactly</u> one positive root. If this root is greater than or equal to one, we take $\mu_2=1$; otherwise, the optimal choice for μ_2 is the value of the root in the open interval $(1/3 + 1/3\gamma_0, 1)$. Although no explicit expression for the root has been found, some numerical work leads us to conjecture that it is a monotonically decreasing function of γ_0. Also,

note that k_0 and k_1 given by (11) and (12) will always lie between 0 and 1 when $\mu_2 \geq \frac{1}{3}$.

The following table gives the optimal choices for μ_2, k_0, and k_1 for some combinations of values of γ_0, γ_1, and γ_2.

$(\gamma_0,\gamma_1,\gamma_2)$	optimal μ_2	optimal k_0	optimal k_1
(1, 1, 1)	.947	.454	.486
(1, 1, 3)	.815	.455	.449
(1, 3, 1)	1.000	.444	.750
(1, 3, 3)	1.000	.400	.750
(3, 1, 1)	.576	.737	.366
(3, 1, 3)	.498	.737	.332
(3, 3, 1)	.748	.713	.692
(3, 3, 3)	.597	.710	.642

On-going research is concerned with multivariate problems, with methods of estimating $\underset{\sim}{K}$ from available data, with certain admissibility requirements, with the evaluation of several design criteria, and with various prior distribution considerations (such as the specification of a prior for σ^2 and the effects of incorrect choices for priors).

REFERENCES

1. BLIGHT, B. J. N. (1971). Some general results on reduced mean square error estimation. The Amer. Statist. 25 (3), 24-25.

2. BOX, G. E. P. and DRAPER, N. R. (1959). A basis for the selection of a response surface design. J. Amer. Statist. Assoc. 54, 622-654.

3. BOX, G. E. P. and DRAPER, N. R. (1963). The choice of a second order rotatable design. Biometrika 50, 335-352.

4. BOX, G. E. P. and WILSON, K. B. (1951). On the experimental attainment of optimum conditions. J. Roy. Statist. Soc. (B) 13, 1-45.

5. DAVID, H. A. and ARENS, B. E. (1959). Optimal spacing in regression analysis. Ann. Math. Statist. 30, 1072-1081.

6. DRAPER, N. E. and LAWRENCE, W. E. (1965). Designs which minimize model inadequacies: cuboidal regions of interest. Biometrika 52, 111-118.

7. JEFFREYS, HAROLD (1961). Theory of Probability. Oxford University Press, London.

8. KARSON, M. J., MANSON, A. R., and HADER, R. J. (1969). Minimum bias estimation and experimental design for response surfaces. Technometrics 11, 461-475.

9. KORTS, D. C. (1971). Estimators based on a closeness criterion over a restricted parameter space for the general linear model of full rank. Univ. of Nor. Car. Instit. of Statist. Mimeo Ser. No. 741.

10. KUPPER, L. L. (1972). Some further comments on mean square error estimation. Submitted to The Amer. Statist.

11. MACDUFFEE, C. C. (1954). Theory of Equations. Wiley, New York.

12. MARKOWITZ, E. (1968). Minimum mean square error estimation of the standard deviation of the normal distribution. The Amer. Statist. 22(3), 26.

13. STEIN, C. (1964). Inadmissibility of the normal estimator for the variance of a normal distribution with unknown mean. Ann. Inst. Statist. Math. 16, 155.

14. STUART, A. (1969). Reduced mean square error estimation of σ^p in normal samples. The Amer. Statist. 23(4), 27.