A NEW ESTIMATION PROCEDURE

FOR

RESPONSE SURFACE MODELS

by

Linda Joy Catron Malone

Dissertation submitted to the Graduate Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

Dr. Raymond H. Myers, Chairman

Dr. Walter H. Carter

Dr. Richard G. Krutchkoff

Dr. Klaus Hinkelmann

Dr. Boyd Harshbarger

Dr. Albert W. Sherdon

ACKNOWLEDGMENTS

It is impossible to thank everyone who has made my stay at Virginia Polytechnic Institute and State University a pleasant one. I would like, however, to mention a few to whom I am especially indebted. First, I would like to express my sincere appreciation to Dr. Raymond Myers for his interest, effort, and encouragement in directing this research and advising me throughout the pursuit of this degree. Also, I am grateful to Dr. W. Hans Carter of the Department of Biostatistics, Medical College of Virginia, Richmond, Virginia, for his suggestions and help in directing this thesis. I would like to thank Dr. Richard Krutchkoff for co-reading this work and for his constructive comments. I would like to extend my gratitude to Dr. Klaus Hinkelmann from whom I had the majority of my classes and Dr. Jesse Arnold who has made my instructorship in the Department an enjoyable experience. I would like to give a special word of thanks to Dr. Boyd Harshbarger for his interest and advice. In addition, I am indebted to Miss Judy Galliher for her efficiency in the preparation of this manuscript.

I would be remiss if I failed to mention my parents. My mother, Gwendolyn Catron, and my father, Byron Catron, encouraged my throughout my life and had such a strong belief in me that I'm sure it was much easier for me to achieve than others. Surely without them the road would have been harder if not impossible to follow. Finally, I wish to thank my husband, Frank, for his patience and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

CHAPTER I

Introduction

Response surface methodology is concerned with techniques for estimating some feature in a scientific or engineering process usually called a response. The response is influenced by a large number of usually continuous quantative input variables which are controlled by the experimenter. Examples of these variables may be time, amount of chemical reactant, pressure, quantity of vitamins, etc. The response is commonly denoted by

$$\eta = f(\xi_1,\xi_2,\xi_3,\ldots,\xi_k;\beta_1,\beta_2,\ldots,\beta_r)$$

where the true functional form f of the response η is unknown and perhaps complicated, the ξ_i are the independent input variables, and the β_i are unknown parameters. It is usually assumed that the experimenter is interested in the response in some operability region 0 and is conducting the experiment inside the operability region in some experimental region R.

It is convenient to standardize or scale the natural variables ξ_i to new variables x_i so that the center of interest of the natural variables becomes the origin of the experimental region R. The geometric configuration of R is usually simple, normally a hypersphere with unit radius or a hypercube with $|x_i| \leq 1$ for all i.

Often the functional form of η is assumed to be a low order polynomial, many times linear or quadratic. That is, η can be represented in terms of the design variables as

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \tag{1.1}$$

or

$$\eta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i}\sum_{\substack{j \\ i<j}} \beta_{ij} x_i x_j \ . \tag{1.2}$$

When little curvature in f is present (i.e., perhaps the experimenter wishes to study η in narrow regions of the x_i), model (1.1) is usually used. Otherwise (1.2) may be more appropriate. The rationale for the polynomial approximation of f is based on the Taylor Series expansion of f around the origin of the experimental region R.

Various criteria have been suggested which could be desirable for an estimator of η. At first most of these were variance reducing criteria although as early as 1941 Hotelling suggested that model inadequacy should be considered. Box and Draper (1959) unified the considerations of variance and bias in considering integrated mean square error averaged over the region of interest R. Of course, this as well as all other properties suggested are influenced by choice of design and/or estimation procedure employed. This thesis attempts to consider three current leading proposals for estimation of the response η and to improve upon the latest approach suggested to make it more usable and more intuitive for the experimenter. The three techniques to be studied were suggested by Box and Draper; Karson, Manson, and Hader; and Kupper and Meydrech, respectively. After presentation of these three approaches, a modified estimator is presented, discussed, and studied. It is then examined in relation to the three proposals already presented.

CHAPTER II

REVIEW OF THE LITERATURE

2.0 Introduction

This survey of the literature will be restricted primarily to recent articles which influenced this particular thesis and the notation generally accepted as standard in response surface methodology and will not attempt to be a history of response surface methodology. The review will be presented in two major sections. The first section will deal with some approaches taken in estimating a response when one is concerned with mean square error. The second section will deal with the same question but with the most recent approach, namely, shrinkage estimators, and give some review of articles which influenced consideration of shrinkage estimators in response surface methodology.

2.1 Mean Square Error Criteria

Until recently (1958) most experimenters considered some sort of variance criteria as useful in response surface analysis. The notable exception to this group was Hotelling who in 1941 investigated the problem of maximizing a response

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots$$

in some region of interest. He fitted a quadratic equation

$$\hat{y}(x) = b_0 + b_1 x + b_2 x^2$$

using the least-squares procedure and after maximizing the response considered the problem of model inadequacy of the quadratic approximation to the true response. Hotelling let B_3 be the coefficient of β_3 and B_4 be the coefficient of β_4 in the expression

$$E(b_1 - \beta_1) = B_3\beta_3 + B_4\beta_4 + \ldots$$

and assuming the variance of b_1 is fixed, he demonstrated how to arrange the design points to make B_3 vanish and to minimize B_4. This is the earliest work concerned with model inadequacy in experimental design. In fact, it wasn't until much later that model inadequacy was again considered in response surface estimation.

Box and Wilson published a paper in 1951 in which they were concerned with finding the combination of factors yielding maximum response in some region R. They also concerned themselves with the bias that could arise if one fits an inadequate polynomial model. To reflect the bias in the least squares estimates of the unknown parameters, they introduced the so-called "alias" matrix.

David and Arens (1959) used mean square error type criteria to find optimal spacing of a single controllable factor when fitting a straight line to a true quadratic response. The criteria introduced were (a) minimum expected squared error integrated over [-1,1] and (b) minimum maximum expected squared error.

Box and Draper were responsible for directing attention toward bias type considerations. Since their paper in 1959,much of the work done in estimation of a response has considered the effect of bias due to model inadequacy.

The problem faced by experimenters was to estimate a response

$$\eta = f(\xi_1,\ldots,\xi_p;\Theta_1,\ldots,\Theta_r)$$

where the functional form $f(\xi,\Theta)$ is unknown, but it is assumed that it can be approximated by a low order polynomial. The coefficients of the polynomial are then the unknown parameters to be estimated. Suppose $\eta(x)$ is a polynomial of degree $d + k - 1$.

$$\eta(x) = \underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2$$

where

$$\underline{x}_1' = [1, x^2, x^3, \ldots, x^{d-1}] \qquad \underline{x}_2' = [x^d, x^{d+1}, \ldots, x^{d+k-1}]$$

$$\underline{\beta}_1' = [\beta_0, \beta_1, \ldots, \beta_{d-1}] \qquad \underline{\beta}_2' = [\beta_d, \beta_{d+1}, \ldots, \beta_{d+k-1}]$$

but for some reason the experimenter has fit a polynomial of degree $d - 1$

$$y(\underline{x}) = X_1\underline{\beta}_1 + \underline{\varepsilon}$$

where $y(\underline{x})$ is a vector of N measured responses corresponding to N different combinations of treatment levels (not all necessarily unique). Ordinary least squares estimation is used to find $\underline{b}_1$

$$\underline{b}_1 = (X_1'X_1)^{-1} X_1'\underline{y}$$

where

$$X_1 = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{d-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{d-1} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & x_N & x_N^2 & & x_N^{d-1} \end{pmatrix} . \quad (X_1 \text{ is } N\times d)$$

Box and Draper (1959) were naturally interested in having $y(\underline{x})$ fit $\eta(\underline{x})$ as well as possible within the experimental region R. They considered the problem of choosing designs which minimized J, the expected mean square error averaged over R. They assumed

$$E(\varepsilon_i) = 0,\ E(\varepsilon_i^2) = \sigma^2 \text{ and } E(\varepsilon_i\ \varepsilon_j) = 0 \text{ for } i \neq j\ .$$

Thus,

$$J = \frac{NK}{\sigma^2} \int_R E\{\hat{y}(x) - \eta(x)\}^2\ dx$$

where

$$K^{-1} = \int_R dx$$

and σ^2 is the experimental error variance. J can be partitioned into two useful components, the integrated variance averaged over the region R and the integrated squared bias averaged over the region R.

$$J = V + B$$

where

$$V = \frac{NK}{\sigma^2} \int_R \text{var}\ \hat{y}(x)\ dx$$

and

$$B = \frac{NK}{\sigma^2} \int_R \{\text{Bias}\ \hat{y}(x)\}^2\ dx.$$

Minimization of J depends on the unknown parameters $\underline{\beta}_2$ and the relative magnitudes of the V and B contributions. However, Box and Draper discovered that unless the variance contribution was very much larger than the bias contribution that designs which minimize J were very close to designs which minimize B alone. These designs, which minimize B alone, are sometimes referred to as all-bias designs.

In order to minimize B, equate the moments of the design up to order $d_1 + d_2$ (where d_1 is the degree $\hat{y}(x)$ and d_2 is the degree of $\eta(x)$) to the corresponding moments of a uniform distribution over R. Let X_2 be the matrix representing bias type terms

$$X_2 = \begin{pmatrix} x_1^d & x_1^{d+1} & \cdots & x_1^{d+k-1} \\ x_2^d & x_2^{d+1} & \cdots & x_2^{d+k-1} \\ \vdots & \vdots & & \vdots \\ x_N^d & x_N^{d+1} & & x_N^{d+k-1} \end{pmatrix}$$

then, in order to minimize B alone the design should satisfy

$$A = \mu_1^{-1}\mu_2$$

where A is the "alias" matrix

$$A = (X_1'X_1)^{-1} X_1'X_2$$

and

$$\mu_1 = K \int_R \underline{x}_1\underline{x}_1' \, d\underline{x} \qquad \mu_2 = K \int_R \underline{x}_1\underline{x}_2' \, d\underline{x} \ .$$

Another approach, this one estimation oriented instead of design oriented, was proposed by Karson, Hader, and Manson in 1967. They dropped the restriction of using least squares estimators. They assumed

$$E(\underline{\varepsilon}) = 0 \qquad \mathrm{Var}\ (\underline{\varepsilon}) = \sigma^2 I$$

$$\eta(\underline{x}) = \underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2$$

where $\underline{x}_1$, $\underline{x}_2$, $\underline{\beta}_1$, $\underline{\beta}_2$ are the same as before and that

$$\hat{y}(\underline{x}) = \underline{x}_1'\underline{b}_1$$

where $\underline{b}_1$ is to be determined. Let $\underline{b}_1$ be some linear function of the observations

$$\underline{b}_1 = T'\underline{y}$$

and

$$\hat{y}(\underline{x}) = \underline{x}_1' \; T'\underline{y}$$

which would yield

$$T' = (X_1'X_1)^{-1} \; X_1'$$

for least squares estimation of $\underline{\beta}_1$.

Considering the conclusion by Box and Draper that bias seemed to be the most important part of the J criterion, Karson, Hader, and Manson decided to minimize B by choice of T and then subject to having minimum B, minimize V. Then

$$\text{Bias} = E[\hat{y}(\underline{x}) - \eta(\underline{x})]$$

$$= \underline{x}_1' \; E(\underline{b}_1) - \eta(\underline{x})$$

$$= \underline{x}_1' \; T' \; (X_1\underline{\beta}_1 + X_2\underline{\beta}_2) - \eta(\underline{x})$$

and

$$(\text{Bias})^2 = \{E(\hat{y}(\underline{x}) - \eta(\underline{x})\}^2$$

$$= [\underline{x}_1'T'(X_1\underline{\beta}_1 + X_2\underline{\beta}_2) - (\underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2)]^2$$

$$= [\underline{x}_1'(T'X_1\underline{\beta}_1 + T'X_2'\underline{\beta}_2 - \underline{\beta}_1) - \underline{x}_2'\underline{\beta}_2]^2 \; .$$

Define $\qquad \underline{\alpha} = \underline{\alpha}(T) = T'X_1\underline{\beta}_1 + T'X_2\underline{\beta}_2 - \underline{\beta}_1$

so that $\qquad (\text{Bias})^2 = (\underline{x}_1'\underline{\alpha} - \underline{x}_2'\underline{\beta}_2)'(\underline{x}_1'\underline{\alpha} - \underline{x}_2'\underline{\beta}_2)$

$$= \underline{\alpha}'\underline{x}_1\underline{x}_1'\underline{\alpha} - 2\underline{\alpha}'\underline{x}_1\underline{x}_2\underline{\beta}_2 + \underline{\beta}_2'\underline{x}_2\underline{x}_2'\underline{\beta}_2 \; .$$

The matrix T was chosen so as to minimize the above "averaged" over the region of interest R. Now

$$B = K \int_R (\text{Bias})^2 \, d\underline{x}$$

and substituting we obtain

$$B = K \int_R (\underline{\alpha}'\underline{x}_1\underline{x}_1'\underline{\alpha} - 2\underline{\alpha}'\underline{x}_1\underline{x}_2'\underline{\beta}_2 + \underline{\beta}_2'\underline{x}_2\underline{x}_2'\underline{\beta}_2) \, d\underline{x}$$

then

$$B = \underline{\alpha}'W_1\underline{\alpha} - 2\underline{\alpha}W_2\underline{\beta}_2 + \underline{\beta}_2'W_3\underline{\beta}_2$$

where

$$W_1 = K \int_R \underline{x}_1\underline{x}_1' \, d\underline{x}$$

$$W_2 = K \int_R \underline{x}_1\underline{x}_2' \, d\underline{x}$$

$$W_3 = K \int_R \underline{x}_2\underline{x}_2' \, d\underline{x} \, .$$

Note that these are constants independent of T. W_1 is symmetric and positive definite so

$$B = (\underline{\alpha} - W_1^{-1}W_2\underline{\beta}_2)' \, W_1(\underline{\alpha} - W_1^{-1}W_2\underline{\beta}_2) + \underline{\beta}_2'(W_3 - W_2'W_1^{-1}W_2)\underline{\beta}_2 \, .$$

Hence in order to minimize B, choose

$$\underline{\alpha} = W_1^{-1}W_2 \, \underline{\beta}_2$$

(since the last of the three terms of B is independent of T). Using the original definition of $\underline{\alpha}$

$$T'X_1\underline{\beta}_1 + T'X_2\underline{\beta}_2 = W_1^{-1}W_2\underline{\beta}_2 + \underline{\beta}_1 \, .$$

Thus, sufficient conditions on T for satisfying the above, independent of $\underline{\beta}_1$ and $\underline{\beta}_2$, and therefore for minimizing B are

$$T'X_1 = I_d$$

and

$$T'X_2 = W_1^{-1}W_2 \ .$$

These can be written in the form

$$T'X = A$$

where

$$X = (X_1 \vdots X_2) \quad \text{and} \quad A = (I_d \vdots W_1^{-1}W_2) \ .$$

Subject to having $T'X = A$ Karson, Hader, and Manson now try to minimize the V component of the J criterion. Now if

$$\hat{y}(\underline{x}) = \underline{x}_1'\underline{b}_1 = \underline{x}_1'T'\underline{y}$$

then

$$\begin{aligned} \text{var } \hat{y}(\underline{x}) &= \underline{x}_1'T'T\underline{x}_1 \ \sigma^2 \\ &= E[\underline{x}_1'T'\underline{y} - \underline{x}_1'T'X\underline{\beta}]^2 \end{aligned}$$

where

$$\underline{\beta}' = (\underline{\beta}_1' \vdots \underline{\beta}_2') \ .$$

Since $E(\underline{b}_1) = A\underline{\beta}$ and since $\underline{b}_1 = T'\underline{y}$ it is known by the Gauss-Markov theorem that var $\hat{y}(X \vdots T)$ will be minimized with respect to T if $\underline{b}_1 = T'\underline{y}$ is the minimum variance unbiased estimator of $A\underline{\beta}$ in a model which includes all the parameters. Therefore

$$\underline{b}_1 = T'\underline{y} = A(X'X)^{-1} X'\underline{y}$$

so that the T matrix which minimizes var $\hat{y}(\underline{x})$ for any $\underline{x}$ is given by

$$T' = A(X'X)^{-1} X'$$

and

$$(\min V|\min B)\ \hat{y}(\underline{x}) = \underline{x}_1' A(X'X)^{-1} X'\underline{y}\ .$$

The above estimator of $\eta(\underline{x})$ requires $(X'X)$ to be nonsingular; recall $X = (X_1 \vdots X_2)$ so in order to use this estimator one needs $N > d + k - 1$ which is the degree of the true model $\eta(\underline{x})$. Hence, all the coefficients in $\eta(\underline{x})$ must be estimated even though the model $\hat{y}(\underline{x})$ is only of order $d - 1 < d + k - 1$.

Parish, Manson, and Hader (1973) and Cote, Manson, and Hader (1973) used the same philosophy and analogous procedures to obtain analogous results as Karson, Hader, and Manson. In the case of Cote, Manson, and Hader the true model $\eta(x)$ was assumed to be a ratio of polynomials while the fitted model $\hat{y}(x)$ was assumed to be a low order polynomial. Parish, Manson, and Hader assumed the true model to be of the form

$$\eta(x) = \alpha + \beta^* e^{\gamma^*\xi} \qquad -\infty < \xi_1 \leq \xi \leq \xi_2 < \infty$$

where α, β^* and γ^* are unknown parameters and ξ is a controllable independent variable. Again they supposed that the model fitted was a low order polynomial.

2.2 Reduction of Mean Square Error Using Shrinkage Estimation

According to Thompson (1968a), Gauss proposed as early as 1809 that the size of the mean square error of an estimator should be used to evaluate the worth of the estimator. Stein (1964), Stuart (1969), Markowitz (1968), Thompson (1968a and 1968b) and others have shown that there exist constants which when multiplied by usual (unbiased or maximum likelihood) estimators for the mean and powers of the standard deviation for a normal distribution reduce the size of the mean square error of the estimator.

Blight (1971) approached the problem of reducing the mean square error of an estimator in a more general way and gave sufficient conditions for uniform improvement over unbiased estimation. Blight considered t an unbiased estimator of τ (where τ is a function of an unknown parameter Θ), and let $V = V(\Theta)$ be the variance of t. He then considered estimators of the form proposed by Stein, et al., namely a constant k times the unbiased estimator. The mean square error of kt is then given by $k^2V + (1-k)^2\tau^2$ and the value of k which minimizes this is

$$k = \frac{\tau^2}{V+\tau^2} \quad .$$

Since k is a function of the parameter Θ, which is generally unknown, the above value of k is not usable. However, if t is inadmissible, it is often possible to find a value for k, say c, for which the mean square error of ct is uniformly better than the mean square error of t.

Kupper and Meydrech (1972) applied the procedure of Blight to the case of simple linear regression and obtained analogous results. Suppose $y(\underline{x}) = \alpha + \beta\underline{x} + \underline{\varepsilon}$. They made the usual assumption that $\underline{\varepsilon} \sim N(0,\sigma^2)$. They used the usual least squares estimator for α. The linear function b which minimizes the mean square error of b is of the form $k\hat{\beta}$ with $\hat{\beta}$ the usual unbiased least squares estimator of β and

$$k = \frac{\beta^2}{\beta^2+[\frac{1}{N\mu_2}]\sigma^2} = \frac{N\mu_2\beta^2}{N\mu_2\beta^2+\sigma^2}$$

where

$$\mu_2 = \frac{1}{N} \sum_{u=1}^{N} x_{i,u}^2 .$$

Motivated by the works cited above Kupper and Meydrech (1972) decided to try using shrinkage estimators in the problem of estimation of a response. They supposed the true model to be

$$\eta(\underline{x}) = \underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2$$

where

$$\underline{x}_1' = (1, x_1, \ldots, x_p) \qquad \underline{x}_2' = (x_1^2, x_2^2, \ldots, x_p^2, x_1x_2, x_1x_3, \ldots, x_{p-1}x_p)$$

$$\underline{\beta}_1' = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$$

$$\underline{\beta}_2' = (\beta_{11}, \beta_{22}, \beta_{33}, \ldots, \beta_{pp}, \beta_{12}, \beta_{13}, \ldots, \beta_{p-1,p})$$

and then fitted the model

$$\hat{y}(\underline{x}) = k_0\hat{\beta}_0 + k_1\hat{\beta}_1x_1 + k_2\hat{\beta}_2x_2 + \ldots + k_p\hat{\beta}_px_p$$

$$\hat{y}(\underline{x}) = \underline{x}_1' K\hat{\underline{\beta}}_1$$

where

$$K = \mathrm{diag}(k_0, k_1, \ldots, k_p)$$

$$\hat{\underline{\beta}}_1 = (X_1'X_1)^{-1}X_1'\underline{y}$$

$$X_1' = (\underline{x}_{11}, \ldots, \underline{x}_{1u}, \ldots, \underline{x}_{1N}) \quad X_1' \text{ is } (p+1)\times N$$

$$\underline{x}_{1u}' = (1, x_{1u}, x_{2u}, \ldots, x_{pu}) .$$

Kupper and Meydrech computed the integrated mean square error of $\hat{y}(\underline{x})$ and then wanted to minimize

$$J = \frac{N\Omega}{\sigma^2} \int_R \mathrm{MSE}(\hat{y}(\underline{x}))$$

where $$\frac{1}{\Omega} = \int_R d\underline{x} \ .$$

It can be easily shown that J can be partitioned into two pieces, a bias component and a variance component. That is;

$$J = V + B$$

where

$$V = \frac{N\Omega}{\sigma^2} \int_R \text{var}(\hat{y}(\underline{x}))\ d\underline{x}$$

and

$$B = \frac{N\Omega}{\sigma^2} \int_R \text{Bias}^2(\hat{y}(\underline{x}))\ d\underline{x} \ .$$

For the particular case in question

$$V = N\Omega \int_R \underline{x}_1'\ K(X_1'X_1)^{-1}\ K\ \underline{x}_1\ d\underline{x}$$

$$= \text{tr}\{K\ C^{-1}\ K\ \Lambda_{11}\}$$

where "tr" denotes the trace operator and

$$C = \frac{1}{N} X_1'X_1$$

and

$$\Lambda_{11} = \Omega \int_R \underline{x}_1\underline{x}_1'\ d\underline{x} \ .$$

It is assumed that the integration is performed elementwise with respect to the entries in the matrix $\underline{x}_1\underline{x}_1'$. Similarly, for the bias component Kupper and Meydrech found

$$B = \frac{N\Omega}{\sigma^2} \int_R [\underline{x}_1'K(\underline{\beta}_1+A\underline{\beta}_2) - (\underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2)]^2\ d\underline{x}$$

so that

$$B = (K\underline{\alpha}_1 + KA\underline{\alpha}_2 - \underline{\alpha}_1)'\ \Lambda_{11}\ (K\underline{\alpha}_1 + KA\underline{\alpha}_2 - \underline{\alpha}_1)$$

$$- 2\underline{\alpha}_2'\ \Lambda_{21}(K\underline{\alpha}_1 + KA\underline{\alpha}_2 - \underline{\alpha}_1) + \underline{\alpha}_2'\Lambda_{22}\underline{\alpha}_2$$

where

$$A = (X_1'X_1)^{-1} X_1'X_2$$

is the $(p+1) \times \frac{p(p+1)}{2}$ alias matrix and

$$\Lambda_{21} = \Omega \int_R \underline{x}_2\underline{x}_1' \, d\underline{x}$$

$$\Lambda_{22} = \Omega \int_R \underline{x}_2\underline{x}_2' \, d\underline{x}$$

$$\underline{\alpha}_1 = \frac{\sqrt{N}}{\sigma} \underline{\beta}_1$$

$$\underline{\alpha}_2 = \frac{\sqrt{N}}{\sigma} \underline{\beta}_2 \quad .$$

Kupper and Meydrech (1972) then showed that for the cuboidal region of interest where $\Omega^{-1} = 2^p$

$$V = k_0^2 + \frac{1}{3} \sum_{i=1}^{p} c^{ii} k_i^2$$

and

$$B = \alpha_0^2(k_0 - 1)^2 + \frac{1}{3} \sum_{i=1}^{p} \alpha_i^2(k_i - 1)^2 + 2k_0(k_0 - 1)\alpha_0$$

$$\text{times} \sum_{i=1}^{p} \sum_{j=1}^{p} \mu_{ij}\alpha_{ij} - \frac{2}{3} (k_0 - 1)\alpha_0 \sum_{i=1}^{p} \alpha_{ii} +$$

$$[k_0 \sum_{i=1}^{p} \sum_{j=1}^{p} \mu_{ij}\alpha_{ij} - \frac{1}{3} \sum_{i=1}^{p} \alpha_{ii}]^2 + \frac{2}{3} \sum_{l=1}^{p} k_l(k_l - 1) \alpha_l \text{ times}$$

$$\sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_{ij} \sum_{h=1}^{p} c^{gh}\mu_{hij} + \frac{1}{3} \sum_{g=1}^{p} [k_g \sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_{ij} \sum_{h=1}^{p} c^{gh}\mu_{hij}]^2$$

$$+ \frac{4}{45} \sum_{i=1}^{p} \alpha_{ii}^2 + \frac{1}{9} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \alpha_{ij}^2$$

where c^{ij} denotes the i,jth element of C^{-1}.

In order to simplify the problem Kupper and Meydrech chose $k_0 = 1$. They also made the following assumptions $\mu_{hij} = 0$, $\mu_{ij} = 0$ and $\mu_{ii} = \mu_2$ for all h, i, and j. The above then simplifies to

$$J = 1 + \frac{1}{3\mu_2}\sum_{i=1}^{p} k_i^2 + \frac{1}{3}\sum_{i=1}^{p}\alpha_i^2(k_i-1)^2 + (\mu_2 - \tfrac{1}{3})^2\left[\sum_{i=1}^{p}\alpha_{ii}\right]^2$$

$$+ \frac{4}{45}\sum_{i=1}^{p}\alpha_{ii}^2 + \frac{1}{9}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\alpha_{ij}^2 .$$

It is then easy to show that in order to minimize J

$$k_i = \frac{\mu_2\alpha_i^2}{1+\mu_2\alpha_i^2} \qquad i = 1,2,\ldots,p$$

or rewriting (using $\alpha_i^2 = \frac{N\beta_i^2}{\sigma^2}$)

$$k_i = \frac{N\mu_2\beta_i^2}{\sigma^2+N\mu_2\beta_i^2}$$

which, it should be noted, is exactly of the form we found for the case of simple linear regression minimizing mean square error of b a linear function of the usual estimator $\hat{\beta} = (X_1'X_1)^{-1} X_1'\underline{y}$.

Kupper and Meydrech compared their J with the J which would be obtained using least squares estimation with the same moment conditions cited earlier and a cuboidal region of interest. Since Draper and Lawrence (1965) worked with least squares estimation and cuboidal regions of interest, Kupper and Meydrech refer to the J obtained in this fashion as J_{dl}. The difference between the Kupper Meydrech J, J_{km}, and J_{dl} is

$$J_{km} - J_{dl} = \frac{1}{3}\sum_{i=1}^{p}[(k_i^2 - 1)/\mu_2 + \alpha_i^2(k_i - 1)^2] .$$

Because of the structure of this difference, each term can be considered separately. Thus, a term in the difference will be zero if $k_i = 1$ and will be negative if

$$\frac{\mu_2\alpha_i^2-1}{1+\mu_2\alpha_i^2} < k_i < 1 .$$

Unfortunately, the α_i^2 involve the unknown parameters β_i. They supposed, however, that a constant M_i can be specified for which it is known that $\alpha_i^2 < M_i$. Because of the monotonicity in α_i^2 of the lower bound of k_i given above,

$$\frac{\mu_2\alpha_i^2-1}{\mu_2\alpha_i^2+1} < \frac{\mu_2 M_i-1}{\mu_2\ M_i+1} .$$

so that any k_i satisfying

$$\frac{\mu_2\alpha_i^2-1}{\mu_2\alpha_i^2+1} < k_i < 1$$

will also satisfy

$$\frac{\mu_2 M_i-1}{\mu_2 M_i+1} \leq k_i \leq \frac{\mu_2 M_i}{1+\mu_2 M_i} .$$

Hence if an upper bound can be specified for at least one of the α_i^2, then they can choose all other $k_i = 1$ and have $J_{km} - J_{d1} < 0$. Since any k_i in the interval above will give $J_{km} - J_{d1} < 0$ some flexibility is left. Kupper and Meydrech chose

$$k_i = \frac{\mu_2 M_i}{1+\mu_2 M_i}$$

because this is the value of k_i that minimizes the maximum of J over the restricted parameter space $\{(\alpha_1^2,\alpha_2^2,\ldots,\alpha_p^2)\colon\ \alpha_i^2 < M_i \text{ for } i = 1,\ldots,p\}$.

CHAPTER III

A NEW SHRINKAGE PROCEDURE FOR RESPONSE SURFACES

3.0 Proposal of a New Procedure

This chapter proposes a new estimation procedure for response surfaces. Various properties of the estimator are derived and discussed.

We saw, in the previous chapter, that Kupper and Meydrech employed shrinkage procedures as a means of reducing integrated mean square error of a fitted response $\hat{y}(\underline{x})$ averaged over the region of interest. They showed that when

$$\hat{y}(\underline{x}) = \hat{\beta}_0 + k_1\hat{\beta}_1 x_1 + k_2\hat{\beta}_2 x_2 + \dots + k_p\hat{\beta}_p x_p$$

and

$$\eta(\underline{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \dots + \beta_{pp} x_p^2$$

$$+ \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \dots + \beta_{p-1,p} x_{p-1} x_p$$

and making the usual assumptions; i.e.,

(1) $\mu_{ijk} = 0$ for all i,j and k; that is, all third moments of the design are chosen to be zero

(2) $\mu_{ij} = 0$ for all $i \neq j$; that is, all mixed second moments of the design are chosen to be zero

the corresponding expression for J is

$$J = 1 + \frac{1}{3\mu_2}\sum_{i=1}^{p} k_i^2 + \frac{1}{3}\sum_{i=1}^{p}\alpha_i^2(k_i - 1)^2 + (\mu_2 - \frac{1}{3})^2 \left[\sum_{i=1}^{p}\alpha_{ii}\right]^2$$

$$+ \frac{4}{45}\sum_{i=1}^{p}\alpha_{ii}^2 + \frac{1}{9}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\alpha_{ij}^2$$

which is minimized when

$$k_i = \frac{\mu_2\alpha_i^2}{1+\mu_2\alpha_i^2} \qquad i = 1,2,\ldots,p$$

where

$$\alpha_i^2 = \frac{N\beta_i^2}{\sigma^2} \qquad i = 1,2,\ldots,p \ .$$

Since k_i is a function of $\alpha_i^2 = \frac{N\beta_i^2}{\sigma^2}$, k_i cannot be used; therefore Kupper and Meydrech put an upper bound on α_i^2, denoted by M_i, and choose k_i to be

$$k_i = \frac{\mu_2 M_i}{1+\mu_2 M_i} \ .$$

It would appear, however, that in many situations it would be unlikely that researchers would be able to put a realistic upper bound on α_i^2.

Kupper and Meydrech noted themselves (1973) that in order for their procedure to have a smaller J value than least squares estimation using the same design [and a cuboidal region of interest], α_i^2 must be less than $2M_i$. Thus M_i may not be taken to be too small. In addition, if M_i is chosen to be too large, the value of k_i is closer to 1 than it should be. As k_i approaches 1,the Kupper and Meydrech method will approach least squares estimation. The optimal design for Kupper and

Meydrech is chosen based on the values of the upper bounds. Thus for upper bounds that are very much larger than α_i^2, they will have a smaller J value than least squares estimation using the same design. However if we can design the experiment and choose a minimum bias design, we could in fact achieve a smaller J than that of Kupper and Meydrech.

Kupper and Meydrech (1972) did show, however, that the value of the shrinkage J was less than the J obtained for least squares estimation (where both used the same designs) for any k_i satisfying

$$\frac{\mu_2\alpha_i^2-1}{\mu_2\alpha_i^2+1} < k_i < 1 .$$

In particular, this result would hold if the optimal design for least squares estimation were chosen. These optimal designs are obtained by finding the moments that minimize J if the parameter values are known. Thus, theoretically the procedure proposed by Kupper and Meydrech is worth considering. The major difficulty associated with its use is choosing an upper bound, M_i, for α_i^2 so that their $k_i = \frac{\mu_2 M_i}{1+\mu_2 M_i}$ will be in the required interval

$$\left(\frac{\mu_2\alpha_i^2-1}{\mu_2\alpha_i^2+1} \ , \ 1\right) .$$

We can see that in order for k_i to be in the interval

$$\frac{\mu_2 M_i}{1+\mu_2 M_i} > \frac{\mu_2\alpha_i^2-1}{\mu_2\alpha_i^2+1}$$

or

$$\mu_2 M_i(\mu_2\alpha_i^2 + 1) > (1+\mu_2)(\mu_2\alpha_i^2 - 1)$$

and so

$$M_i > \frac{\alpha_i^2}{2} - \frac{1}{2\mu_2} \quad .$$

Since $\mu_2 > 0$, $M_i > \frac{\alpha_i^2}{2} - \frac{1}{2\mu_2}$ is implied by $M_i > \frac{\alpha_i^2}{2}$. Therefore, as long as $M_i > \frac{\alpha_i^2}{2}$ (or $2M_i > \alpha_i^2$), Kupper and Meydrech's choice for k_i will be in the interval, and their resulting J will be smaller than the J obtained using least squares estimation withe the same μ_2.

The ideas of Kupper and Meydrech are interesting. The work by Blight (1971) can be generalized to more than simple linear regression, Suppose we have

$$\hat{y}(\underline{x}) = X\, K\hat{\underline{\beta}}$$

where

$$\begin{aligned} \hat{\underline{y}}(\underline{x}) &\text{ is } N\times 1 \\ X &\text{ is } N\times p \\ K &\text{ is } p\times p \text{ diagonal} \\ \underline{\beta} &\text{ is } p\times 1 \quad . \end{aligned}$$

Consider minimizing with respect to k_i a quantity ψ where

$$\psi = \operatorname{tr} E\{(K\hat{\underline{\beta}} - \underline{\beta})(K\hat{\underline{\beta}} - \underline{\beta})'\} = \sum_{i=1}^{p} E(k_i b_i - \beta_i)^2 \quad . \tag{3.1}$$

As we can see, ψ is the sum of the mean square errors of each individual "shrunken" regression coefficient. From (3.1)

$$\psi = \text{tr } E\{K\underline{bb}'K' - K\underline{b\beta}' - \underline{\beta b}K' - \underline{\beta\beta}'\}$$

where $\underline{b} = \hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$

$$\psi = \text{tr } E\{K(X'X)^{-1}X'\underline{yy}'X(X'X)^{-1}K' - K(X'X)^{-1}X'\underline{y\beta}'$$

$$- \underline{\beta y}'X(X'X)^{-1}K' - \underline{\beta\beta}'\} .$$

Substituting for $\underline{y}$ and simplifying

$$\psi = \text{tr } \{K\underline{\beta\beta}'K' + K(X'X)^{-1}K'\sigma^2 - K\underline{\beta\beta}' - \underline{\beta\beta}'K - \underline{\beta\beta}'\} .$$

A typical diagonal element of the square matrix whose trace we will eventually take is of the following form

$$k_i^2\beta_i^2 + \frac{1}{N\mu_2} k_i^2 \sigma^2 - 2k_i \beta_i^2 - \beta_i^2 .$$

In order to minimize the trace with respect to each k_i, we take the derivative with respect to each k_i and equate to zero and solve. Hence to minimize ψ we would need to have

$$k_i = \frac{\beta_i^2}{\left(\beta_i^2 + \frac{\sigma^2}{N\mu_2}\right)} = \frac{N\mu_2\beta_i^2}{\sigma^2 + N\mu_2\beta_i^2} \qquad i = 1,2,3,\ldots,p.$$

By virtue of the above discussion, the concept of shrinkage estimation in multiple regression does seem to have some theoretical foundation; and in particular, the specific choice of the shrinkage factor k_i to be

$$k_i = \frac{N\mu_2\beta_i^2}{\sigma^2+N\mu_2\beta_i^2} \qquad i = 1,2,\ldots,p$$

does seem to have justification. We therefore propose an alternative shrinkage procedure. It seems that instead of guessing an upper bound for α_i^2, we should logically try to estimate k_i from the sample data. The normal estimator we would consider is substituting into the formula for k_i the least squares estimates of the parameters of which k_i is a function. Hence, use

$$k_i = \frac{N\mu_2\hat{\beta}_i^2}{\hat{\sigma}^2+N\mu_2\hat{\beta}_i^2}$$

where

$$\underline{\hat{\beta}} = (X'X\ )^{-1}X'\underline{y}$$

and
$$\hat{\sigma}^2 = (\underline{y}'\underline{y} - \underline{\hat{\beta}}'X_1'y)/(N-p) \ .$$

The resulting estimator $\hat{y}(\underline{x})$ of $\eta(\underline{x})$ would then be

$$\hat{y}(\underline{x}) = \hat{k}_0\hat{\beta}_0 + \hat{k}_1\hat{\beta}_1x_1 + \hat{k}_2\hat{\beta}_2x_2 + \ldots + \hat{k}_p\hat{\beta}_px_p \ .$$

3.1 Distribution of k

Since we are estimating k, it is a random variable with a distribution of its own. Let us write k as

$$\hat{k}_i = \frac{\mu_2\hat{\alpha}_i^2}{1+\mu_2\hat{\alpha}_i^2} = \frac{1}{1+\dfrac{1}{\mu_2\hat{\alpha}_i^2}}$$

where $\hat{\alpha}_i^2 = \frac{N\hat{\beta}_i^2}{\hat{\sigma}^2}$. Suppose we make the usual assumptions on the errors. Now $\hat{\beta}_i \sim N(\beta_i, \sigma^2/N\mu_2)$ implies that $\frac{\hat{\beta}_i}{\sigma/\sqrt{N\mu_2}} \sim N(\beta_i\sqrt{N\mu_2}/\sigma, 1)$. Thus

$\frac{\hat{\beta}_i^2 N\mu_2}{\sigma^2} \sim \chi'^2_{1,\lambda}$ where the noncentrality parameter $\lambda = \frac{1}{2\sigma^2} N\mu_2\beta_i^2$. We also know that $\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{N-p}}{N-p}$ independent of $\hat{\beta}_i$. Therefore $\frac{1}{\mu_2\hat{\alpha}_i^2} = \frac{[\hat{\sigma}^2/\sigma^2]}{\left(\frac{N\mu_2\hat{\beta}_i^2}{\sigma^2}\right)} \sim$

$\frac{\chi^2_{N-p}/(N-p)}{\chi'^2_{1,\lambda}} = \frac{1}{F'_{1,N-p,\lambda}}$. Let $x \sim F'_{1,N-p,\lambda}$ and we have that k_i is distributed as $\frac{1}{\left(1+\frac{1}{x}\right)} = \frac{x}{x+1}$ $\quad 0 \le x < \infty$. Therefore, the density of $\hat{k}_i$ is the same as the density of $\frac{x}{x+1}$. Solving for x we get $x = \frac{\hat{k}}{1-\hat{k}}$ $\quad 0 \le k < 1$. Note that $\frac{dx}{d\hat{k}} = \left(\frac{1}{1-\hat{k}}\right)^2$. Using the form of the density for a noncentral F distribution with N-p degrees of freedom and noncentrality parameter λ, the Jacobian of the transformation from x to k, and a well-known theorem on transformation of variables we obtain

$$g(\hat{k}_1) = \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{2i+N-1}{2}\right)\left(\frac{1}{N-p}\right)^{\frac{2i+1}{2}} \lambda^i e^{-\lambda} \left(\frac{\hat{k}}{1-\hat{k}}\right)^{\frac{2i-1}{2}}}{\Gamma\left(\frac{N-p}{2}\right)\left(\frac{2i+1}{2}\right) i!\left(1+\frac{\hat{k}}{(N-p)(1-\hat{k})}\right)^{\frac{2i+N-p+1}{2}} (1-\hat{k})^2}.$$

This distribution is used in Section 4.1 in comparing $\hat{k}$ to Kupper and Meydrech's k. A computer program using the first fifteen terms of the summation when integrated from 0 to .99 gives .9966 of the density. (Note the density is not defined at $\hat{k} = 1$.)

3.2 Distribution of $\hat{y}(x)$

In order to compute J we must know the distribution of $\hat{y}(\underline{x})$. Since this seemed rather difficult to accomplish for the general case, we looked at the easiest possible situation, namely

$$\hat{y}(x) = \hat{\beta}_0 + \hat{k}_1\hat{\beta}_1 x$$

$$\eta(x) = \beta_0 + \beta_1 x + \beta_{11}x^2$$

and considered the distribution of $\hat{k}_1\hat{\beta}_1$, where

$$\hat{k}_1 = \frac{N\mu_2\hat{\beta}_1^2}{\hat{\sigma}^2 + N\mu_2\hat{\beta}_1^2} \quad .$$

We know that

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2/N\mu_2)$$

and

$$\hat{\sigma}^2 \sim \frac{\sigma^2\chi^2_{N-2}}{N-2}$$

and that $\hat{\beta}_1$ is independent of $\hat{\sigma}^2$. Let $x = \hat{\beta}_1$ then

$$f(x) = \frac{\sqrt{N\mu_2}}{\sigma\sqrt{2\pi}} \; e^{\frac{-N\mu_2}{2\sigma^2}(x-\beta_1)^2} \quad .$$

Similarly, let $y = \hat{\sigma}^2$, then we have

$$g(y) = \left(\frac{N-2}{\sigma^2}\right)^{\frac{N-2}{2}} \frac{1}{\Gamma(\frac{N-2}{2})2^{\frac{N-2}{2}}} \; y^{(\frac{N-2}{2})-1} \; e^{-\left(\frac{N-2}{\sigma^2}\right)(\frac{y}{2})} \quad .$$

Since x and y are independent, we have the joint distribution of x and y

$$h(x,y) = f(x)\cdot g(y)$$

and so

$$h(x,y) = \frac{\sqrt{N\mu_2}}{\sigma\sqrt{2\pi}} \left(\frac{N-2}{\sigma^2}\right)^{\frac{N-2}{2}} \frac{1}{\Gamma(\frac{N-2}{2})2^{\frac{N-2}{2}}} \; e^{\frac{-N\mu_2}{2\sigma^2}(x-\beta_1)^2} \quad \text{times}$$

$$y^{(\frac{N-2}{2})-1} \; e^{-\left(\frac{N-2}{2\sigma^2}\right)y} \quad .$$

Now we can write

$$z = \hat{k}_1\hat{\beta}_1 \quad .$$

Let $a = N\mu_2$ and using $x = \hat{\beta}_1$ we get

$$z = \frac{ax^3}{y+ax^2} \quad .$$

Define a variable $t = y+ax^2$. Now let us find the joint distribution of z and t. Using the definition of t and z we have

$$z = \frac{ax^3}{y+ax^2} = \frac{ax^3}{t} \Rightarrow x = (\frac{zt}{a})^{\frac{1}{3}}$$

$$t = y + ax^2 \Rightarrow y = t - ax^2$$

substituting in for x we get

$$y = t - a(\frac{zt}{a})^{\frac{2}{3}} .$$

We know that y must be non-negative so we must require

$$t - a(\frac{zt}{a})^{\frac{2}{3}} \geq 0$$

or

$$t^{\frac{2}{3}}[t^{\frac{1}{3}} - a(\frac{z}{a})^{\frac{2}{3}}] \geq 0 .$$

From the definition of t, we know it is always positive. We also observe that $t^{\frac{2}{3}} > 0$. Therefore we must have

$$t^{\frac{1}{3}} - a(\frac{z}{a})^{\frac{2}{3}} > 0$$

or

$$t^{\frac{1}{3}} > a(\frac{z}{a})^{\frac{2}{3}}$$

which implies

$$t > \frac{a^3 z^2}{a^2} = az^2 .$$

To compute the Jacobian of the transformation we need the following derivatives:

$$\frac{\partial x}{\partial z} = \frac{1}{3}\left(\frac{t}{a}\right)^{\frac{1}{3}} z^{-\frac{2}{3}}$$

$$\frac{\partial x}{\partial t} = \frac{1}{3}\left(\frac{z}{a}\right)^{\frac{1}{3}} t^{-\frac{2}{3}}$$

$$\frac{\partial y}{\partial z} = -\frac{2}{3}a\left(\frac{t}{a}\right)^{\frac{2}{3}} z^{-\frac{1}{3}}$$

$$\frac{\partial y}{\partial t} = 1 - \frac{2}{3}a\left(\frac{z}{a}\right)^{\frac{2}{3}} t^{-\frac{1}{3}}$$

so that the

$$\text{Jacobian} = \begin{vmatrix} \frac{1}{3}\left(\frac{t}{a}\right)^{\frac{1}{3}} z^{-\frac{2}{3}} & \frac{1}{3}\left(\frac{z}{a}\right)^{\frac{1}{3}} t^{-\frac{2}{3}} \\ -\frac{2}{3}a\left(\frac{t}{a}\right)^{\frac{2}{3}} z^{-\frac{1}{3}} & 1 - \frac{2}{3}a\left(\frac{z}{a}\right)^{\frac{2}{3}} t^{-\frac{1}{3}} \end{vmatrix}$$

and hence

$$\text{Jacobian} = \frac{1}{3}\left(\frac{t}{a}\right)^{\frac{1}{3}} z^{-\frac{2}{3}} .$$

We can observe that J is always non-negative. Thus the density of z can be written as

$$\ell(z) = \int_t m(t,z)\, dz\, dt$$

$$= \int_{az^2}^{\infty} \frac{\sqrt{N\mu_2}}{\sigma\sqrt{2\pi}} \left(\frac{N-2}{\sigma^2}\right)^{\frac{N-2}{2}} \frac{1}{\Gamma\left(\frac{N-2}{2}\right) 2^{\frac{N-2}{2}}}\, e^{-\frac{N\mu_2}{2\sigma^2}\left[\left(\frac{zt}{a}\right)^{\frac{1}{3}}-\beta_1\right]^2}$$

$$\text{times} \left(t-a\left(\frac{zt}{a}\right)^{\frac{2}{3}}\right)^{\frac{N-2}{2}-1} e^{-\left[\frac{N-2}{2\sigma^2}\right]\left[t-a\left(\frac{zt}{a}\right)^{\frac{2}{3}}\right]}$$

$$\text{times}\ \frac{1}{3}\left(\frac{t}{a}\right)^{\frac{1}{3}} z^{-\frac{2}{3}}\, dt\ .$$

It appears that it would be difficult to use this distribution to find

$$B = \frac{NK}{\sigma^2} \int_R [\text{Bias } \hat{y}(x)]^2 \quad \text{and} \quad V = \frac{NK}{\sigma^2} \int_R \text{var } \hat{y}(x)$$

(where $K^{-1} = \int_R dx$). Also we should again note that β_0 was assumed to be zero and we considered only the case of simple linear regression.

3.3 An Approximation for J

We would like to compare the integrated mean square error, J, for our proposed procedure with other procedures. Namely, we are interested in comparing with all bias design, all variance design, and an arbitrary design for least squares estimation. We are also interested in comparing the optimum J values for least squares and shrinkage estimation with each other and with our J. The optimum values are unattainable in practice (since we find the optimal design knowing the parameters) but we are interested in seeing how close our procedure might come to the optimal value. It is also interesting to see how much improvement the shrinkage procedure will give over the least squares procedure. We might also like to see if changing $\hat{k}_1$ slightly will reduce our J. Since the distribution of $\hat{y}(\underline{x})$, even in the most elementary case, appears hard to use, and since we want to compare J values for multivariable situations, another approach seems necessary. We will use the Taylor Series expansion to get an approximation of J.

Let us consider a specific elementary situation and we will be able to see how the structure of J generalizes for any number of variables upon completion. The case of simple linear regression where

$$\eta(\underline{x}) = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2$$

$$\hat{y}(x) = \hat{k}_0 \hat{\beta}_0 + \hat{k}_1 \hat{\beta}_1 x_1$$

is too specialized to help us in determining a general expression. Therefore let us consider the case where

$$\eta(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 +$$

$$\beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

or written in vector notation

$$\eta(\underline{x}) = X_1 \underline{\beta}_1 + X_2 \underline{\beta}_2$$

and

$$\hat{y}(\underline{x}) = b_0 + \hat{k}_1 b_1 x_1 + \hat{k}_2 b_2 x_2 + \hat{k}_3 b_3 x_3$$

or $$\hat{y}(\underline{x}) = X\hat{K}\hat{\underline{\beta}}$$

where

$$\underline{\beta}_1' = [\beta_0, \beta_1, \beta_2, \beta_3]$$

$$\underline{\beta}_2' = [\beta_{11}, \beta_{22}, \beta_{33}, \beta_{12}, \beta_{13}, \beta_{23}]$$

$$\hat{\underline{\beta}}_1 = (X_1'X_1)^{-1} X_1'\underline{y}$$

$$X_1 = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} \\ 1 & x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1N} & x_{2N} & x_{3N} \end{pmatrix} \qquad X_2 = \begin{pmatrix} x_{11}^2 & x_{21}^2 & x_{31}^2 & x_{11}x_{21} & x_{11}x_{31} & x_{21}x_{31} \\ x_{12}^2 & x_{22}^2 & x_{32}^2 & x_{12}x_{22} & x_{12}x_{32} & x_{22}x_{32} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1N}^2 & x_{2N}^2 & x_{3N}^2 & x_{1N}^2x_{2N} & x_{1N}x_{3N} & x_{2N}x_{3N} \end{pmatrix}$$

and

$$\hat{K} = \begin{pmatrix} 1 & & & 0 \\ & \hat{k}_1 & & \\ & & \hat{k}_2 & \\ 0 & & & \hat{k}_3 \end{pmatrix} .$$

We are considering the case where k_0, the coefficient of b_0, is taken to be 1 as it was for Kupper and Meydrech's (1972) work. It is easy to show that the bias of $\hat{y}(\underline{x})$ is

$$\begin{aligned} \text{Bias } \hat{y}(\underline{x}) &= E(\hat{y}(\underline{x})) - \eta(\underline{x}) \\ &= \mu_2(\beta_{11} + \beta_{22} + \beta_{33}) + x_1E\hat{k}_1b_1 \\ &\quad + x_2E\hat{k}_2b_2 + x_3E\hat{k}_3b_3 - \beta_1x_1 - \beta_2x_2 - \beta_3x_3 \\ &\quad - \beta_{11}x_1^2 - \beta_{22}x_2^2 - \beta_{33}x_3^2 - \beta_{12}x_1x_2 - \beta_{13}x_1x_3 - \beta_{23}x_2x_3 . \end{aligned}$$

Similarly we can find the variance of $\hat{y}(\underline{x})$

$$\begin{aligned}
\text{var } \hat{y}(\underline{x}) &= E(\hat{y}(\underline{x}))^2 - [E(\hat{y}(\underline{x}))]^2 \\
&= [\beta_0 + \mu_2[\beta_{11} + \beta_{22} + \beta_{33}]]^2 + x_1^2 E\hat{k}_1^2 b_1^2 \\
&+ x_2^2\, E\hat{k}_2^2 b_2^2 + x_3^2 E\hat{k}_3^2 b_3^2 + 2x_1 Eb_0 \hat{k}_1 b_1 \\
&+ 2x_2 Eb_0 \hat{k}_2 b_2 + 2x_3 Eb_0 \hat{k}_3 b_3 + 2x_1 x_2 E\hat{k}_1 \hat{k}_2 b_1 b_2 \\
&+ 2x_1 x_3 E\hat{k}_1 \hat{k}_3 b_1 b_3 + 2x_2 x_3 E\hat{k}_2 \hat{k}_3 b_2 b_3 \\
&- [\beta_0 + \mu_2(\beta_{11} + \beta_{22} + \beta_{33}) + x_1 E\hat{k}_1 b_1 + x_2 E\hat{k}_2 b_2 \\
&+ x_3 E\hat{k}_3 b_3]^2 \ .
\end{aligned}$$

In order to evaluate the expected values contained in the expressions for the bias and variance of our estimated response, we will use a Taylor Series expansion around the means of the parameters β_0, β_1, β_2, β_3, and σ^2. The Taylor Series expansion of a function f(x,y) about the point (a,b) can be written as follows

$$\begin{aligned}
f(x,y) &= f(a,b) + (x-a)\frac{\partial}{\partial x} f(x,y)\Big|_{(a,b)} + (y-b)\frac{\partial}{\partial y} f(x,y)\Big|_{(a,b)} \\
&+ \frac{1}{2}(x-a)^2 \frac{\partial^2}{\partial x^2} f(x,y)\Big|_{(a,b)} + (x-a)(y-b)\frac{\partial^2}{\partial x\,\partial y} f(x,y)\Big|_{(a,b)} \\
&+ \frac{1}{2}(y-b)^2 \frac{\partial^2}{\partial y^2} f(x,y)\Big|_{(a,b)} \ .
\end{aligned}$$

Because we are expanding around the means of the respective parameters, all first order terms and mixed second order terms will go to zero when we apply the expected value operator.

The following values will also be needed in order to approximate J;

$$\text{var } b_0 = \sigma^2/N$$

$$\text{var } b_i = \sigma^2/N\mu_2 \qquad i = 1,2,3$$

and

$$\text{var } \hat{\sigma}^2 = 2\sigma^4/(N-4).$$

The computation of individual pieces of the variance and bias expressions follow. They are then assembled to give the final results for the three variable case. Let

$$f(b_1,\hat{\sigma}^2) = \hat{k}_1 b_1 = \frac{b_1^3\mu_2 N}{\hat{\sigma}^2+\mu_2 N b_1^2} = \frac{b_1^3\mu_2 N}{\hat{\psi}_1}$$

where

$$\hat{\psi}_i = \hat{\sigma}^2 + \mu_2 N b_i^2 .$$

Then

$$E(\hat{k}_1 b_1) \tilde{=} \frac{\beta_1^3\mu_2 N}{\psi_1} + \frac{1}{2}\text{ var } b_1 \cdot \frac{\partial^2}{\partial b_1^2} f(b_1,\hat{\sigma}^2) +$$

$$\frac{1}{2}\text{ var } \hat{\sigma}^2 \cdot \frac{\partial^2 f(b_1,\hat{\sigma}^2)}{\partial(\hat{\sigma}^2)^2}$$

differentiating

$$\frac{\partial f(b_1,\hat{\sigma}^2)}{\partial b_1} = \frac{3\mu_2 N b_1^2\hat{\sigma}^2 + \mu_2^2 N^2 b_1^4}{\hat{\psi}_1^2}$$

and taking second derivatives we have

$$\frac{\partial^2 f(b_1,\hat{\sigma}^2)}{\partial b_1^2} = \frac{2\mu_2 N b_1 \hat{\sigma}^2 W_1}{\hat{\psi}_1^3} .$$

where

$$\hat{W}_i = 3\hat{\sigma}^2 - \mu_2 N b_i^2 .$$

Also we have

$$\frac{\partial f(b_1,\hat{\sigma}^2)}{\partial\hat{\sigma}^2} = \frac{-b_1^3\mu_2 N}{\hat{\psi}_1^2}$$

and

$$\frac{\partial^2 f(b_1,\hat{\sigma}^2)}{\partial(\hat{\sigma}^2)^2} = \frac{2b_1^3\mu_2 N}{\hat{\psi}_1^3} .$$

Thus

$$E(\hat{k}_1 b_1) \simeq \frac{\beta_1^3\mu_2 N}{\psi_1} + \frac{\beta_1\sigma^4 W_1}{\psi_1^3} + \frac{2\beta_1^3\mu_2\sigma^4 N}{\psi_1^3(N-4)} .$$

Because of symmetry, we have similarly for $E(\hat{k}_2 b_2)$ and $E(\hat{k}_3 b_3)$

$$E(k_2 b_2) \simeq \frac{\beta_2^3\mu_2 N}{\psi_2} + \frac{\beta_2\sigma^4 W_2}{\psi_2^3} + \frac{2\beta_2^3\mu_2\sigma^4 N}{\psi_2^3(N-4)}$$

and

$$E(k_3 b_3) \simeq \frac{\beta_3^3\mu_2 N}{\psi_3} + \frac{\beta_3\sigma^4 W_3}{\psi_3^3} + \frac{2\beta_3^3\mu_2\sigma^4 N}{\psi_3^3(N-4)} .$$

Hence the bias of $\hat{y}(\underline{x})$ is

$$\text{Bias } \hat{y}(\underline{x}) \simeq \mu_2(\beta_{11} + \beta_{22} + \beta_{33}) + \sum_{i=1}^{3} x_i \left[\frac{\beta_i^3\mu_2 N}{\psi_i} + \frac{\beta_i\sigma^4 W_i}{\psi_i^3} + \frac{2\beta_i^3\mu_2\sigma^4 N}{\psi_i^3(N-4)}\right] - \sum_{i=1}^{3} \beta_i x_i - \sum_{i=1}^{3} x_{ii}^2 - \sum_{i<j=1}^{2} \sum^{3} \beta_{ij} x_i x_j .$$

Squaring the bias and dropping all terms that go to zero upon integration, we obtain

$$[\text{Bias } \hat{y}(\underline{x})]^2 \cong [\mu_2 \sum_{i=1}^{3} \beta_{ii}]^2 + [\sum_{i=1}^{3} x_i \left(\frac{\beta_i \mu_2 N}{\psi_i} + \frac{\beta_i \sigma^4 W_i}{\psi_i^3} + \frac{2\beta_i^3 \mu_2 \sigma^4 N}{\psi_i^3 (N-4)} \right)]^2 + (\Sigma \beta_i x_i)^2 + (\Sigma \beta_{ii} x_i^2)^2$$

$$+ (\Sigma \sum_{i<j=1}^{3} \beta_{ij} x_i x_j)^2 - 2\mu_2 \sum_{i=1}^{3} \beta_{ii} \sum_{i=1}^{3} \beta_{ii} x_i^2$$

$$- 2 \sum_{i=1}^{3} x_i \left(\frac{\beta_i \mu_2 N}{\psi_i} + \frac{\beta_i \sigma^4 W_i}{\psi_i^3} + \frac{2\beta_i^3 \mu_2 N \sigma^4}{\psi_i^3 (N-4)} \right) \sum_{i=1}^{3} \beta_i x_i \, .$$

Let

$$\Phi_i = \frac{\beta_i^3 \mu_2 N}{\psi_i} + \frac{\beta_i \sigma^4 W_i}{\psi_i^3} + \frac{2\beta_i^3 \mu_2 \sigma^4 N}{\psi_i^3 (N-4)}$$

from which we obtain that

$$[\text{Bias } \hat{y}(\underline{x})]^2 \cong [\mu_2 \sum_{i=1}^{3} \beta_{ii}]^2 + \sum_{i=1}^{3} x_i^2 \Phi_i^2 + \sum_{i=1}^{3} \beta_i^2 x_i^2$$

$$+ \sum_{i=1}^{3} \beta_{ii}^2 x_i^4 + 2 \Sigma \sum_{i<j=1}^{3} \beta_{ii} \beta_{jj} x_i^2 x_j^2$$

$$+ 2 \Sigma \sum_{i<j=1}^{3} \beta_{ij}^2 x_i^2 x_j^2 - 2\mu_2 \sum_{i=1}^{3} \beta_{ii} \Sigma \beta_{ii} x_i^2$$

$$- 2 \sum_{i=1}^{3} \beta_i x_i \sum_{i=1}^{3} x_i \Phi_i \, .$$

Integrating and averaging over the region of interest we get

$$B \doteq \frac{N}{\sigma^2}[\mu_2 \sum_{i=1}^{3} \beta_{ii}]^2 + \frac{N}{3\sigma^2}\sum_{i=1}^{3}\Phi_i^2 + \frac{N}{3\sigma^2}\sum_{i=1}^{3}\beta_i^2$$

$$+ \frac{N}{5\sigma^2}\sum_{i=1}^{3}\beta_{ii}^2 + \frac{2N}{9\sigma^2}\sum\sum_{i<j=1}^{3}\beta_{ii}\beta_{jj} + \frac{N}{9\sigma^2}\sum\sum_{i<j=1}^{3}\beta_{ij}^2$$

$$- \frac{2N\mu_2}{3\sigma^2}\sum_{i=1}^{3}\sum_{j=1}^{3}\beta_{ii}\beta_{jj} - \frac{2N}{3\sigma^2}\sum_{i=1}^{3}\beta_i\Phi_i \ .$$

Now let us turn our attention to the variance component of the integrated mean squared error. A number of computations are necessary for the Taylor Series expansion. These are reported in the following pages. First, consider

$$[E\ \hat{y}(\underline{x})]^2 \doteq \{\beta_0 + \mu_2(\beta_{11} + \beta_{22} + \beta_{33}) + \sum_{i=1}^{3} x_i \left(\frac{\beta_i^3\mu_2 N}{\psi_i} + \frac{\beta_i\sigma^4 W_i}{\psi_i^3} + \frac{2\beta_i^3\mu_2\sigma^4 N}{\psi_i^3(N-4)} \right) \}^2$$

$$= [\beta_0 + \mu_2(\sum_{i=1}^{3}\beta_{ii})]^2 + \sum_{i=1}^{3} x_i^2 \left(\frac{\beta_i^3\mu_2 N}{\psi_i} + \frac{\beta_i\sigma^4 W_i}{\psi_i^3} + \frac{2\beta_i^3\mu_2\sigma^4 N}{\psi_i^3(N-4)} \right)^2 + \text{terms involving}$$

$$x_i x_j \ (i \neq j) + \text{terms involving } x_i \ . \qquad (3.2)$$

Note that all terms involving $x_i x_j$ $(i \neq j)$ and terms involving x_i

will go to zero upon integration. Hence, we need only consider the terms written out in (3.2). Now for $[\hat{y}(\underline{x})]^2$ we have

$$[\hat{y}(\underline{x})]^2 = [\beta_0 + \mu_2 \sum_{i=1}^{3} \beta_{ii}]^2 + \sum_{i=1}^{3} k_i^2 b_i^2 x_i^2 + \text{terms involving}$$

$$x_i + \text{terms involving } x_i x_j \quad i \neq j \ .$$

The following derivatives will be needed:

first,

$$\frac{\partial[\hat{y}(x)]^2}{\partial b_0} = 2(\hat{\beta}_0 + \mu_2 \sum_{i=1}^{3} \hat{\beta}_{ii}) + 2 \sum_{i=1}^{3} \hat{\beta}_i^2 x_i^2 \hat{k}_i^2$$

and

$$\frac{\partial^2[\hat{y}(x)]^2}{\partial b_0^2} = 2$$

second,

$$\frac{\partial[\hat{y}(x)]^2}{\partial \hat{\sigma}^2} = \sum_{i=1}^{3} \left(\frac{-2\mu_2^2 N^2 \hat{\beta}_i^6}{\hat{\psi}_i^3} \right) x_i^2$$

and

$$\frac{\partial^2[\hat{y}(x)]^2}{\partial(\hat{\sigma}^2)^2} = \sum_{i=1}^{3} \left(\frac{6\mu_2^2 N^2 \hat{\beta}_i^6}{\hat{\psi}_i^4} \right) x_i^2$$

third,

$$\frac{\partial[\hat{y}(x)]^2}{\partial b_i} = \sum_{i=1}^{3} \left(\frac{6\hat{\sigma}^2 \hat{\beta}_i^5 N^2 \mu_2^2 + 2\hat{\beta}_i^7 N^3 \mu_2^3}{\hat{\psi}_i^3} \right) x_i^2$$

and

$$\frac{\partial^2[\hat{y}(x)]^2}{\partial b_i^2} = \sum_{i=1}^{3} \left(\frac{30\hat{\beta}_i^4 \mu_2^2 N^2 \sigma^4 + 8\beta_i^6 \mu_2^3 N^3 \sigma^2 + 2\beta_i^8 \mu_2^4 N^4}{\hat{\psi}_i^4} \right) x_i^2 \ .$$

Combining the pieces (i.e. $E(\hat{y}(\underline{x}))^2 - [E\hat{y}(\underline{x})]^2$), integrating and averaging over the region of interest, we obtain

$$V \cong \frac{N}{3\sigma^2} \sum_{i=1}^{3} \left(\frac{\beta_i^6 \mu_2^2 N^2}{\psi_i^2} \right) + 1 + \frac{N\sigma^2}{N-4} \sum_{i=1}^{3} \frac{2\mu_2^2 N^2 \beta_i^6}{\psi_i^4}$$

$$+ \frac{1}{6\mu_2} \sum_{i=1}^{3} \frac{30\beta_i^4 \mu_2^2 N^2 \sigma^4 + 8\beta_i^6 \mu_2^3 N^3 \sigma^2 + 2\beta_i^8 \mu_2^4 N^4}{\psi_i^4}$$

$$- \frac{N}{3\sigma^2} \sum_{i=1}^{3} \left(\frac{\beta_i^3 \mu_2 N}{\psi_i} + \frac{\beta_i \sigma^4 W_i}{\psi_i^3} + \frac{2\beta_i^3 \mu_2 N \sigma^4}{\psi_i^3 (N-4)} \right)^2$$

where we recall from earlier work that

$$\psi_i = \sigma^2 + \mu_2 N \beta_i^2$$

$$W_i = 3\sigma^2 - \mu_2 N \beta_i^2 \quad .$$

When we put the two expressions for V and B together, we get integrated mean square error J. We can easily see how the three variable case can be generalized to any number of variables, say p. Thus suppose

$$\hat{y}(\underline{x}) = b_0 + \hat{k}_1 b_1 x_1 + \hat{k}_2 b_2 x_2 + \ldots + \hat{k}_p b_p x_p = b_0 + \sum_{i=1}^{p} \hat{k}_i b_i x_i$$

and

$$\eta(\underline{x}) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{\substack{i=1 \\ i<j}}^{p-1} \sum_{j=1}^{p} \beta_{ij} x_i x_j + \sum_{i=1}^{p} \beta_{ii} x_i^2$$

then

$$B \tilde{=} \frac{N}{\sigma^2}[\mu_2 \sum_{i=1}^{p} \beta_{ii}]^2 + \frac{N}{3\sigma^2}\sum_{i=1}^{p}\Phi_i^2 + \frac{N}{3\sigma^2}\sum_{i=1}^{p}\beta_i^2$$

$$+ \frac{N}{5\sigma^2}\sum_{i=1}^{p}\beta_{ii}^2 + \frac{2N}{9\sigma^2}\sum_{i<}^{p-1}\sum_{j}^{p}\beta_{ii}\beta_{jj}$$

$$+ \frac{N}{9\sigma^2}\sum_{i<}^{p-1}\sum_{j=1}^{p}\beta_{ij}^2 - \frac{2N\mu_2}{3\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\beta_{ii}\beta_{jj}$$

$$- \frac{2N}{3\sigma^2}\sum_{i=1}^{p}\beta_i\Phi_i$$

and

$$V \tilde{=} \frac{N}{3\sigma^2}\sum_{i=1}^{p}\left(\frac{\beta_i^6\mu_2^2N^2}{\psi_i^2}\right) + 1 + \left(\frac{N\sigma^2}{N-p-1}\right)\sum_{i=1}^{p}\frac{2\mu_2^2N^2\beta_i^6}{\psi_i^4}$$

$$+ \frac{1}{6\mu_2}\sum_{i=1}^{p}\left(\frac{30\beta_i^4\mu_2^2N^2\sigma^4 + 8\beta_i^6\mu_2^3N^3\sigma^2 + 2\beta_i^8\mu_2^4N^4}{\psi_i^4}\right)$$

$$- \frac{N}{3\sigma^2}\sum_{i=1}^{p}\left(\frac{\beta_i^3\mu_2N}{\psi_i} + \frac{\beta_i\sigma^4W_i}{\psi_i^3} + \frac{2\beta_i^3\mu_2N\sigma^4}{\psi_i^3(N-p-1)}\right)^2 .$$

3.4 A Check on Accuracy of the Taylor Series Expansion of J

Two things are considered in checking the accuracy of the Taylor Series expansion of J. First we consider the expression

$$\psi_i = \hat{\sigma}^2 + \mu_2 N\hat{\beta}_i^2$$

which appears in the denominator of most of the terms of J. When the ratio of the mean of ψ_i to the standard deviation of ψ_i is around 3 the expansion is relatively good. The mean and standard deviation of ψ_i are easily obtained as follows: It is known that the mean and variance of a noncentral chi-square random variable, $\chi'^2_{\nu\lambda}$ can be expressed in terms of degrees of freedom, ν, and noncentrality parameter λ. Let us consider the one variable case for simplicity, then

$$\text{mean } \chi'^2 = \nu + \lambda$$

$$\text{var } \chi'^2 = 2(\nu+\lambda)(1 + \frac{\lambda}{\nu+\lambda}) \ .$$

Now we know that

$$\hat{\sigma}^2 \sim (\chi^2/\nu)\sigma^2$$

so that

$$E(\hat{\sigma}^2) = \sigma^2 \text{ and var } \hat{\sigma}^2 = \frac{\sigma^4}{\nu^2} (2\nu) = \frac{2\sigma^4}{\nu} \ .$$

Also we know that

$$\frac{N\mu_2}{\sigma^2} b_i^2 \sim \chi'^2_{1,\lambda}$$

hence

$$E(\frac{N\mu_2}{\sigma^2} b_1^2) = 1 + \frac{\beta_1^2 N\mu_2}{\sigma^2}$$

and

$$E(N\mu_2 b_1^2) = \sigma^2 + \beta_1^2 N\mu_2 \ .$$

Therefore, the mean of ψ_i is

$$\text{mean } (\hat{\sigma}^2 + \mu_2 \, Nb_1^2) = \sigma^2 + \sigma^2 + \beta_1^2 N\mu_2 = 2\sigma^2 + \beta_1^2 N\mu_2 \ .$$

Now, turning our attention to the variance of ψ_i we have

$$\text{var } (\hat{\sigma}^2 + \mu_2 Nb_1^2) = \text{var } \sigma^2 + \mu_2^2 N^2 \text{ var } b_1^2 \ .$$

Consider

$$\mathrm{var}\left(\frac{N\mu_2}{\sigma^2} b_1^2\right) = 2\left[1 + \frac{\beta_1^2 N\mu_2}{\sigma^2}\right]\left[1 + \frac{\frac{\beta_1^2 N\mu_2}{\sigma^2}}{1+\frac{\beta_1^2 N\mu_2}{\sigma^2}}\right]$$

and so

$$N^2\mu_2^2 \,\mathrm{var}\, b_1^2 = 2\sigma^4\left[1 + \frac{\beta_1^2 N\mu_2}{\sigma^2}\right]\left[1 + \frac{\beta_1^2 N\mu_2}{\sigma^2+\beta_1^2 N\mu_2}\right] .$$

Hence, the variance of ψ_1 is

$$\mathrm{var}\,(\hat{\sigma}^2 + \mu_2 N b_1^2) = \frac{2\sigma^4}{N-2} + 2\sigma^4\left[1 + \frac{\beta_1^2 N\mu_2}{\sigma^2}\right]\left[1 + \frac{\beta_1^2 N\mu_2}{\sigma^2+\beta_1^2 N\mu_2}\right]$$

which can be simplified to

$$\mathrm{var}\,(\hat{\sigma}^2 + \mu_2 N b_1^2) = \frac{2\sigma^4}{N-2} + 2\sigma^4 + 4\beta_1^2\sigma^2 N\mu_2 \quad .$$

Finally we obtain

$$\frac{\text{mean } \psi_1}{\text{standard deviation } \psi_1} = \frac{2\sigma^2+\beta_1^2 N\mu_2}{\left(\frac{2\sigma^4}{N-2} + 2\sigma^4+4\beta_1^2\sigma^2 N\mu_2\right)^{1/2}} \quad .$$

In order to more easily use the above ratio, let us rewrite it in terms of $\alpha_1^2 = \frac{N\beta_1^2}{\sigma^2}$ which yields

$$\frac{\text{mean } \psi_1}{\text{standard deviation } \psi_1} = \frac{2+\alpha_1^2\mu_2}{[2(\frac{N-1}{N-2})+4\alpha_1^2\mu_2]^{1/2}} \quad .$$

We can show, using the ratio written as a function of α_1^2, that for all μ_2 in the range from .3 to 1 the ratio will be larger than three for $\alpha_1^2 > 30$. Therefore, we would feel our Taylor Series expansion is reliable for parameter sets inwhich $\alpha_1^2 > 30$. For the multivariable case the solution is a little more elusive. It appears that as long as $\alpha_i^2 > 30$ for all i, and the number of observations taken is at least twice the number of variables fitted that the Taylor Series expansion is reliable.

Whenever the reliability of the Taylor Series expansion was in question, Monte Carlo studies were used to compare the performance of our estimator with least squares estimation, minimum bias design and minimum variance design. The Monte Carlo work revealed that the procedure itself worked as well, and sometimes better in the regions where the Taylor Series expansion was not reliable,as when the expansion was reliable.

CHAPTER IV

NUMERICAL COMPARISONS

4.0 Introduction

In this chapter we will make numerical comparisons among the procedures using results developed in the previous chapter. It is hoped these examples will help the reader understand and evaluate the proposed procedure. All tables are accurate to thousandths.

In order to do numerical work with the proposed procedure, a large amount of computer work is required. To obtain any numerical results, we need to specify all of the $\underline{\beta}$ vector (both terms fit in the model and higher order terms omitted from the model), σ^2, N, and the second design moment μ_2. Some Monte Carlo is also necessary, the results are reported in section three.

4.1 Closeness Criterion

We decided to investigate how close the estimated $\hat{k}$ is to the optimal k in relation to the distance between the optimal k and the k used by Kupper and Meydrech. Suppose we let the upper bound M_1 on α_1^2 be denoted by

$$M_1 = \delta_1\alpha_1^2 \qquad \delta_1 > 1/2 \ .$$

We should note that we are assuming that the experimenter is able to find an upper bound that will reduce the integrated mean square error below that of least squares estimation. If the experimenter is unable to select an upper bound, the following work would not be applicable.

The closeness criterion adopted is that of

$$\Pr[|\hat{k} - k_{opt}| < |k_{KM} - k_{opt}|] \; . \qquad (4.1)$$

In the above notation

$$\hat{k} = \frac{\mu_2 N \hat{\beta}_1^2}{\hat{\sigma}^2 + \mu_2 N \hat{\beta}_1^2}$$

$$k_{opt} = k_{optimum} = \frac{\mu_2 N \beta_1^2}{\sigma^2 + \mu_2 N \beta_1^2}$$

$$k_{KM} = k_{Kupper\text{-}Meydrech} = \frac{\mu_2 M_1}{1 + \mu_2 M_1} \; .$$

We observe that for $\delta_1 > 1$ we have $k_{KM} > k_{opt}$ and for $\frac{1}{2} < \delta_1 < 1$ we have $k_{KM} < k_{opt}$. Rewriting (4.1) results in

$$\Pr[2k_{opt} - k_{KM} < \hat{k} < k_{KM}] \qquad \delta_1 > 1 \; (4.2)$$

$$\Pr[k_{KM} < \hat{k} < 2k_{opt} - k_{KM}] \quad \frac{1}{2} < \delta_1 < 1 \; (4.3) \; .$$

Using the density of $\hat{k}$ along with a computer program to evaluate (4.2) and (4.3), we get results reported in the tables on pages 45 and 46.

It can be seen that if the guess at the upper bound is very good our k has a low probability of being closer to the optimal value than the k used by Kupper and Meydrech. However, the upper bound used could be in the acceptable range (i.e. $M_1 > \frac{1}{2}\alpha_1^2$) and be either too close to the lower limit or too much larger than α_1^2, in which case, our k will be closer to the optimal k more often than the Kupper-Meydrech k.

TABLE Ia. $*Pr[|\hat{k} - k_{optimum}| < |k_{KM} - k_{opt}|]$

δ_1	μ_2	N	$\alpha_1^2 = \frac{N\beta_1^2}{\sigma^2}$	*Probability
.55	.4	15	30	.73644
.60	.4	15	30	.65198
.65	.4	15	30	.55156
.75	.4	15	30	.35580
.85	.4	15	30	.19158
.55	.4	20	30	.74880
.60	.4	20	30	.67293
.65	.4	20	30	.57503
.55	.6	20	30	.55316
.60	.6	20	30	.51107
.60	.4	20	40	.59566
.65	.4	20	40	.52331
1.80	.4	15	30	.49643
1.85	.4	15	30	.51238
1.95	.4	15	30	.54126
2.00	.4	15	30	.55433
2.50	.4	15	30	.64806
3.00	.4	15	30	.69959
3.50	.4	15	30	.72961
3.00	.6	15	30	.51722
3.50	.6	15	30	.53458
1.80	.4	20	30	.51933

TABLE Ib. *Pr[$|\hat{k} - k_{optimum}| < |k_{KM} - k_{opt}|$]

δ_1	μ_2	N	$\alpha_1^2 = \frac{N\beta_1^2}{\sigma^2}$	*Probability
1.95	.4	20	30	.56470
2.0	.4	20	30	.57780
2.5	.4	20	30	.66923
3.0	.4	20	30	.71672
3.5	.4	20	30	.74299
2.5	.6	20	30	.50107
3.0	.6	20	30	.52776
3.5	.6	20	30	.54282
2.5	.4	15	40	.56802
3.0	.4	15	40	.60536
3.5	.4	15	40	.62666
1.95	.4	20	40	.50831
2.5	.4	20	40	.58569
3.0	.4	20	40	.61877
2.0	.9	15	10	.55082
1.95	.9	15	10	.53734
1.90	.9	15	10	.52302
1.85	.9	15	10	.50776
1.80	.9	15	10	.49149
1.70	.9	15	10	.45565
1.60	.9	15	10	.41481
1.50	.9	15	10	.36815
1.30	.9	15	10	.25308
1.10	.9	15	10	.09846

4.2 Sensitivity Study on $\hat{k}$

It was suggested that perhaps there might be a way of improving the shrinkage factor $\hat{k}$. One way of changing $\hat{k}$ is to insert the maximum likelihood estimator for σ^2 instead of using the unbiased estimator for σ^2. This is done and the integrated mean square errors are compared. The use of the unbiased $\hat{\sigma}^2$ in $\hat{k}$ produces smaller J values than do the use of the maximum likelihood $\hat{\sigma}^2$ in $\hat{k}$. The insertion of the maximum likelihood estimator of σ^2 makes the corresponding value of $\hat{k}$ larger than when the unbiased estimator is used. Next, we will try making $\hat{k}$ smaller and observe the affect on J. A Monte Carlo study is conducted. The results appear in the table on page 48. All values of J are computed for $\alpha_1^2 = 6$, $\mu_2 = .4$ and $\sigma^2 = 1.0$. We let x denote the point in the interval (-1,1) where the squared bias plus the variance is evaluated. It can be seen from the table there is no justification in choosing a different estimator for $\hat{k}$.

4.3 Monte Carlo Study

As was stated in Chapter III, the Taylor Series expansion of J is not a good approximation to J for all parameter sets; therefore, a Monte Carlo study is done comparing the proposed procedure with least squares estimation minimum bias design. The quantity computed is un-integrated mean square error. Only selected parameter sets are considered as obviously one cannot Monte Carlo all possible ones. The value for mean square error for least squares estimation minimum bias design is not computed by Monte Carlo procedures. The mean square errors reported for our shrinkage estimation are averages of ten sets

TABLE II. Unintegrated M. S. E. for Corresponding $\hat{k}$

N	β_1	β_{11}	x	$\hat{k}^* = .8\hat{k}$	$\hat{k}^* = .85\hat{k}$	$\hat{k}^* = .9\hat{k}$	$\hat{k}^* = .95\hat{k}$	$\hat{k}^* = \hat{k}$
14	.65	.327	.8	.218516	.216289	.215413	.215891	.217717
14	.65	.327	.4	.085009	.086790	.088909	.091365	.094158
14	.65	.327	-.4	.118876	.117488	.116445	.115746	.115392
14	.65	.327	-.8	.151080	.155493	.161273	.168422	.176939
14	.65	.655	.8	.259862	.254432	.250357	.247637	.246270
14	.65	.655	.4	.091123	.094552	.098319	.102424	.106866
14	.65	.655	-.4	.156968	.154102	.151577	.149392	.147549
14	.65	.655	-.8	.141534	.149015	.157860	.168070	.179644
14	.65	1.309	.8	.401743	.390516	.380633	.372098	.364915
14	.65	1.309	.4	.132811	.139424	.146375	.153665	.161294
14	.65	1.309	-.4	.255885	.249716	.243887	.238397	.233246
14	.65	1.309	-.8	.152134	.165927	.181082	.197600	.215480
20	.55	.274	.8	.151923	.150590	.150200	.150761	.152268
20	.55	.274	.4	.061388	.062731	.064313	.066133	.068191
20	.55	.274	-.4	.083880	.082943	.082241	.081774	.081543

of 1000 samples of size N. The standard error is included and all cases with less than two standard errors difference are starred. The parameter α_1^2 is set equal to 6 in all results given. It should be recalled that the Taylor Series expansion for J is not reliable when $\alpha_1^2 < 30$.) The intercept β_0 is chosen to be zero. Once again x denotes the point in the range (-1,1) where the mean square error was evaluated. We can observe from the table that when $\alpha_1^2 = 6$, there are parameter sets where our procedure gives smaller mean square error than least squares estimation using a minimum bias design.

4.4 Comparisons of J Values for the One Variable Case

We would like to make further comparisons in terms of the J criterion by using the Taylor Series expansion derived in Chapter IV. We compute the optimal J for least squares estimation. This value is derived from a knowledge of the parameters and is optimal in the sense that the second design moment used is the one which minimizes J. We use the best k for the shrinkage J, substituting in actual parameter values. Hence we use

$$k = \frac{\mu_2 N\beta_1^2}{\sigma^2+\mu_2 N\beta_1^2} \quad .$$

We use the same second moment for our proposed shrinkage procedure and for the shrinkage estimation using optimal k. Thus the shrinkage J is not computed using the optimal μ_2 as is the case for least squares estimation. We also compute the J for least squares estimation having a minimum bias design and having a minimum variance design. The value of σ^2 used throughout all comparisons is one.

TABLE III. Monte Carlo Study

N	μ_2	β_1	β_{11}	X	Standard Error	Our M.S.E.	Least Squares Min. Bias Design MSE
14	.40	.80	.400892	.4	.0026	.141229	.158716
14	.40	.80	.400892	-.8	.0022	.265389	.314278
14	.40	.80	.801784	-.8	.0025	.269453	.318542
14	.40	.80	1.603567	-.8	.0068	.323211	.335590*
20	.40	.67	.335410	.4	.0009	.102299	.111101
20	.40	.67	.335410	-.8	.0033	.182773	.219995
14	.40	.33	.163663	.4	.0006	.023533	.026453
14	.40	.33	.163663	-.8	.0004	.044231	.052380
14	.40	.33	.327327	-.8	.0004	.044908	.053090
14	.40	.33	.654654	-.8	.0015	.053867	.055933*
20	.40	.27	.136931	.4	.0001	.017044	.018517
14	.40	.73	.365963	.4	.0022	.117701	.132264
14	.40	.73	.365963	-.8	.0019	.221163	.261899
14	.40	.73	.731925	-.8	.0021	.224556	.265452
14	.40	.73	1.463850	-.8	.0056	.269351	.279666*
20	.40	.61	.306186	.4	.0008	.085245	.092585
20	.40	.61	.306186	-.8	.0028	.152310	.183329
14	.40	.57	.283473	.4	.0013	.070613	.079358
14	.40	.57	.283473	-.8	.0012	.132700	.157139
14	.40	.57	.566947	-.8	.0017	.134732	.159271
20	.40	.47	.237171	.4	.0005	.051138	.055551
20	.40	.47	.237171	-.3	.0017	.091383	.109997

A second set of tables will omit the shrinkage J and the optimum J for least squares estimation and include the value of J we will obtain using least squares estimation but the same second design moment as used in our procedure.

We can observe from the tables that for a variety of designs and parameter sets the proposed procedure is an improvement over existing procedures. There are, however, parameter sets for which the performance of our procedure is inferior to least squares estimation using either an all bias or an all variance design. It generally seems to be less sensitive to the parameters than either least squares procedure. Except when the amount of bias present is very large, our procedure will either produce smaller values of J than the two least squares procedures or will be in close competition with the better of the two.

4.5 Comparison of J Values for the Two Variable Case

The Taylor Series expansion is used to compare our procedure for the two variable case with the J value for the optimum design using least squares estimation, the shrinkage J using the optimum design, the J value for least squares estimation using minimum bias design, and the J value for least squares estimation using an all variance design. The optimum designs for the shrinkage estimation and least squares estimation are attained by minimizing J with respect to μ_2 using the actual parameter values. Of course, this can not be done in actuality. These values are included so that we might compare them with each other, and so that it can be seen how far away each of the other procedures is from the best values attainable. We do not

TABLE IVa. Comparison of J Values for One Variable

$\beta_0 = 0 \quad \mu_2 = .38$

Relationship of V to B	N	β_1	β_{11}	Optimum J for L.S.*	Shrinkage J Using μ_2=.38	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
V = B	12	2.0	1.299	3.718	3.67488	3.79920	12.12853	3.85778
V = 2B	12	2.0	.864	2.656	2.64792	2.79680	6.11416	2.83083
V = 4B	12	2.0	.526	2.052	2.13391	2.29508	3.10383	2.31682
V = 2B	18	2.0	.706	2.656	2.66259	2.79680	6.11416	2.78685
V = 4B	18	2.0	.429	2.052	2.14857	2.29508	3.10383	2.27284
V = 6B	18	2.0	.286	1.790	1.98070	2.13122	2.12065	2.10496
V = B	12	4.0	1.299	3.718	3.70861	3.79920	12.12853	3.75804
V = 2B	12	4.0	.864	2.656	2.68166	2.79680	6.11416	2.73109
V = 4B	12	4.0	.526	2.052	2.16764	2.29503	3.10383	2.21702
V = 6B	12	4.0	.351	1.790	1.99977	2.13122	2.12065	2.04911

*L.S. = Least Squares Estimation

TABLE IVb. Comparison of J Values for One Variable

$\beta_0 = 0 \quad \mu_2 = .38$

Relationship of V to B	N	β_1	β_{11}	Optimum J for L.S.*	Shrinkage J Using μ_2=.38	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
V = B	15	4.0	1.162	3.718	3.71095	3.79920	12.12853	3.75036
V = 2B	15	4.0	.773	2.656	2.68400	2.79680	6.11416	2.72337
V = 4B	15	4.0	.470	2.052	2.16999	2.29508	3.10383	2.02942
V = 6B	15	4.0	.314	1.790	2.00211	2.13122	2.12065	2.04153
V = $\frac{1}{2}$B	18	4.0	1.541	5.755	5.76431	5.80192	24.14482	5.79707
V = B	18	4.0	1.060	3.718	3.71253	3.79920	12.12851	3.74524
V = 2B	18	4.0	.706	2.656	2.68557	2.79680	6.11415	2.71829
V = 4B	18	4.0	.429	2.052	2.17156	2.29508	3.10383	2.20427
V = 6B	18	4.0	.286	1.790	2.00368	2.13122	2.12065	2.03645

*L.S. = Least Squares Estimation

TABLE IVc. Comparison of J Values for One Variable

$\beta_0 = 0 \quad \mu_2 = .4$

Relationship of V to B	N	β_1	β_{11}	Optimum J for L.S.*	Shrinkage J Using μ_2=.4	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
V = 2B	12	2.0	.864	2.656	2.62872	2.79680	6.11416	2.79464
V = 4B	12	2.0	.526	2.052	2.10191	2.29508	3.10383	2.26784
V = 6B	12	2.0	.351	1.790	1.92986	2.13122	2.12065	2.09578
V = 2B	15	2.0	.773	2.656	2.63664	2.79680	6.11415	2.77078
V = 4B	15	2.0	.470	2.052	2.10983	2.29508	3.10383	2.24397
V = 6B	15	2.0	.314	1.790	1.93778	2.13122	2.12065	2.07191
V = 2B	18	2.0	.706	2.656	2.64201	2.79680	6.11415	2.75454
V = 4B	18	2.0	.429	2.052	2.11520	2.29508	3.10383	2.22773
V = 6B	18	2.0	.286	1.790	1.94315	2.13122	2.12065	2.05568
V = B	12	4.0	1.299	3.718	3.71178	3.79920	12.12853	3.75655
V = 2B	12	4.0	.864	2.656	2.65926	2.79680	6.11416	2.70396

*L.S. = Least Squares Estimation

TABLE IVd. Comparison of J Values for One Variable

$\beta_0 = 0 \quad \mu_2 = .4$

Relationship of V to B	N	β_1	β_{11}	Optimal J for L.S.*	Shrinkage J Using μ_2=.4	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
V = 4B	12	4.0	.526	2.052	2.13246	2.29508	3.10383	2.17717
V = 6B	12	4.0	.351	1.790	1.96040	2.13122	2.12065	2.00514
V = B	15	4.0	1.162	3.718	3.71390	3.79920	12.12851	3.74957
V = 2B	15	4.0	.773	2.656	2.66138	2.79680	6.11415	2.69706
V = 4B	15	4.0	.470	2.052	2.13458	2.29508	3.10383	2.17029
V = 6B	15	4.0	.314	1.790	1.96252	2.13122	2.12065	1.99870
V = B	18	4.0	1.060	3.718	3.71532	3.79920	12.12851	3.74496
V = 2B	18	4.0	.706	2.656	2.66280	2.79680	6.11415	2.69247
V = 4B	18	4.0	.429	2.052	2.13600	2.29508	3.10383	2.16567
V = 6B	18	4.0	.286	1.790	1.96394	2.13122	2.12065	1.99360

*L.S. = Least Squares Estimation

TABLE IVe. Comparison of J Values for One Variable

$\beta_0 = 0 \quad \mu_2 = .5$

Relationship of V to B	N	β_1	β_{11}	Optimum J for L.S.*	Shrinkage J Using μ_2=.5	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
V = 2B	12	2.0	.864	2.656	2.68580	2.79680	6.11416	2.79408
V = 4B	12	2.0	.526	2.052	2.02730	2.29508	3.10383	2.13558
V = 6B	12	2.0	.351	1.790	1.81223	2.13122	2.12065	1.92051
V = 2B	15	2.0	.773	2.656	2.69096	2.79680	6.11415	2.77811
V = 4B	15	2.0	.470	2.052	2.03246	2.29508	3.10383	2.11961
V = 6B	15	2.0	.314	1.790	1.81739	2.13122	2.12065	1.90454
V = 2B	18	2.0	.706	2.656	2.69445	2.79680	6.11415	2.76737
V = 4B	18	2.0	.429	2.052	2.03594	2.29508	3.10383	2.10886
V = 6B	18	2.0	.286	1.790	1.82087	2.13122	2.12065	1.89378

*L.S. = Least Squares Estimation

TABLE IVf. Comparison of J Values for One Variable

$\beta_0 = 0 \quad \mu_2 = .5$

Relationship of V to B	N	β_1	β_{11}	Optimum J for L.S.*	Shrinkage J Using μ_2=.5	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
V = 2B	12	4.0	.864	2.656	2.70560	2.79680	6.11416	2.73430
V = 4B	12	4.0	.526	2.052	2.04709	2.29508	3.10383	2.07586
V = 6B	12	4.0	.351	1.790	1.83202	2.13122	2.12065	1.86071
V = 2B	15	4.0	.773	2.656	2.70690	2.70680	6.11415	2.72996
V = 4B	15	4.0	.470	2.052	2.04845	2.29508	3.10383	2.07146
V = 6B	15	4.0	.314	1.790	1.83338	2.13122	2.12065	1.85643
V = 2B	18	4.0	.706	2.656	2.70787	2.79680	6.11415	2.72689
V = 4B	18	4.0	.429	2.052	2.04936	2.29508	3.10383	2.06840
V = 6B	18	4.0	.286	1.790	1.83429	2.13122	2.12065	1.85334

*L.S. = Least Squares Estimation

TABLE Va. Comparison of J Values for One Variable

$\beta_0 = 0$

N	β_1	β_{11}	σ^2	μ_2	*L.S. J With Same μ_2	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
15	8	5	16	.3	5.3316	5.0833	14.1667	4.3944
17	8	5	36	.3	4.2847	4.0494	7.9630	3.4772
31	8	5	64	.3	4.3121	4.0764	8.1250	3.4949
15	12	5	16	.3	5.3316	5.0833	14.1667	4.3030
15	12	5	36	.3	4.1597	3.9259	7.2222	3.2225
15	12	5	64	.3	3.7496	3.5208	4.7917	2.9190
15	16	5	16	.3	5.3316	5.0833	14.1667	4.2677
15	16	5	36	.3	4.1596	3.9259	7.2222	3.1515
15	16	5	64	.3	3.7496	3.5208	4.7917	2.8123
15	8	10	16	.3	11.6597	11.3333	51.6666	10.7225
17	8	10	36	.3	7.4722	7.1975	26.8518	6.6647
31	8	10	64	.3	7.5816	7.3056	27.5000	6.7644
15	12	10	16	.3	11.6597	11.3333	51.6666	10.6311
15	12	10	36	.3	6.9722	6.7037	23.8889	6.0350
15	12	10	64	.3	5.3316	5.0833	14.1667	4.5010
15	16	10	16	.3	11.6597	11.3333	51.6666	10.5958
15	16	10	36	.3	6.9722	6.7037	23.8889	5.9440
15	16	10	64	.3	5.3316	5.0833	14.1667	4.3944

TABLE Vb. Comparison of J Values for One Variable

$\beta_0 = 0$

N	β_1	β_{11}	σ^2	μ_2	*L.S. J With Same μ_2	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
15	8	1	16	.4	2.7542	3.0833	2.1667	2.1466
15	12	1	16	.4	2.7542	3.0833	2.1667	2.0929
15	8	1	16	.6	2.2611	3.0833	2.1667	2.1267
17	8	1	36	.6	2.1867	3.0420	1.9185	1.9059
31	8	1	64	.6	2.1886	3.0431	1.9250	1.9095
15	12	1	16	.6	2.2611	3.0833	2.1667	2.1061
15	12	1	36	.6	2.1778	3.0370	1.8889	1.8350
15	12	1	64	.6	2.1486	3.0208	1.7917	1.7651
15	16	1	16	.6	2.2611	3.0833	2.1667	2.0925
15	16	1	36	.6	2.1778	3.0370	1.8889	1.8154
15	16	1	64	.6	2.1486	3.0208	1.7917	1.7329
15	12	1	36	.8	1.9611	3.0370	1.8889	1.8485
15	16	1	64	.8	1.9052	3.0208	1.7917	1.6711
15	2	0	1	.6	2.1111	3.0000	1.6667	1.6017
15	12	1	64	.8	1.9052	3.0208	1.7917	1.6900
15	16	1	36	.8	1.9611	3.0370	1.8889	1.8372

TABLE Vc. Comparison of J Values for One Variable

$\beta_0 = 0$

N	β_1	β_{11}	σ^2	μ_2	*L.S. J With Same μ_2	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
68	2	0	9	.6	2.1111	3.0000	1.6667	1.6392
271	2	0	36	.6	2.1111	3.0000	1.6667	1.2824
15	6	0	1	.6	2.1111	3.0000	1.6667	1.5609
15	6	0	1	.6	2.1111	3.0000	1.6667	1.5582
15	6	0	9	.6	2.1111	3.0000	1.6667	1.6017
30	6	0	9	.6	2.1111	3.0000	1.6667	1.5787
31	6	0	36	.6	2.1111	3.0000	1.6667	1.6382
15	12	0	1	.6	2.1111	3.0000	1.6667	1.5577
15	12	0	9	.6	2.1111	3.0000	1.6667	1.5676
15	12	0	36	.6	2.1111	3.0000	1.6667	1.6017
15	2	0	1	.8	1.8333	3.0000	1.6667	1.4430
68	2	0	9	.8	1.8333	3.0000	1.6667	1.4651
271	2	0	36	.8	1.8333	3.0000	1.6667	1.1829
15	6	0	1	.8	1.8333	3.0000	1.6667	1.4198
15	6	0	9	.8	1.8333	3.0000	1.6667	1.4430
31	6	0	36	.8	1.8333	3.0000	1.6667	1.4645
15	12	0	1	.8	1.8333	3.0000	1.6667	1.4171
15	12	0	1	.8	1.8333	3.0000	1.6667	1.4235

TABLE Vd. Comparison of J Values for One Variable

$\beta_0 = 0$

N	β_1	β_{11}	σ^2	μ_2	*L.S. J With Same μ_2	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
15	12	0	36	.8	1.8333	3.0000	1.6667	1.4430
30	8	1	16	.3	3.3910	3.1667	2.6667	2.3699
30	12	1	16	.3	3.3910	3.1667	2.6667	2.3210
30	8	5	16	.3	7.4410	7.1667	26.6666	6.4199
30	8	5	36	.3	5.0972	4.8519	12.7778	4.1749
30	12	5	16	.3	7.4410	7.1667	26.6666	6.3711
30	12	5	36	.3	5.0972	4.8519	12.7778	4.0762
30	12	5	64	.3	4.2769	4.0417	7.9167	3.3191
30	16	5	16	.3	7.4410	7.1667	26.6666	6.3536
30	16	5	36	.3	5.0972	4.8519	12.7778	4.0378
30	16	5	64	.3	4.2769	4.0417	7.9167	3.2558
30	8	10	16	.3	20.0972	19.6667	101.6666	19.0761
30	8	10	36	.3	10.7222	10.4074	46.1111	9.7999
30	12	10	16	.3	20.0972	19.6667	101.6666	19.0273
30	12	10	36	.3	10.7222	10.4074	46.1111	9.7011
30	12	10	64	.3	7.4410	7.1667	26.6666	6.4832
30	16	10	16	.3	20.0972	19.6667	101.6666	19.0094
30	16	10	36	.3	10.7222	10.4074	46.1111	9.6628
30	16	10	64	.3	7.4410	7.1667	26.6666	6.4198

TABLE Ve. Comparison of J Values for One Variable

$\beta_0 = 0$

N	β_1	β_{11}	σ^2	μ_2	*L.S. J With Same 2	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
30	8	1	16	.4	2.8417	3.1667	2.6667	2.3096
30	12	1	16	.4	2.8417	3.1667	2.6667	2.2817
30	12	1	36	.4	2.7444	3.0741	2.1111	2.0736
30	16	1	16	.4	2.8417	3.1667	2.6667	2.2713
30	16	1	36	.4	2.7444	3.0741	2.1111	2.0515
30	8	1	36	.6	2.2444	3.0741	2.1111	2.0726
30	12	1	36	.6	2.2444	3.0741	2.1111	2.0454
30	12	1	64	.6	2.1861	3.0417	1.9167	1.8583
30	16	1	36	.6	2.2444	3.0741	2.1111	2.0353
30	16	1	64	.6	2.1861	3.0417	1.9167	1.8411
30	12	1	64	.8	1.9771	3.0417	1.9167	1.8958
30	16	1	64	.8	1.9771	3.0417	1.9167	1.8859
100	8	8	9	.3	67.2222	66.2099	380.9255	66.1269
100	12	8	9	.3	67.2222	66.2099	380.9255	66.1178
100	16	8	9	.3	67.2222	66.2099	380.9255	66.1138
50	12	8	9	.3	35.2222	34.6049	191.2963	34.1252
50	16	8	9	.3	35.2222	34.6049	191.2963	34.1179

TABLE VI. Comparison of J Values for Two Variables

N	β_0	β_1	β_2	β_{11}	β_{22}	β_{12}	σ^2	μ_2	Optimum J for L.S.*	Optimum J Shrinkage	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
15	0	4	1	0	0	0	.09	.4	1.66667	1.66664	3.00000	1.66667	2.70546
271	0	4	1	0	0	0	9.0	.4	1.66667	1.66651	3.00000	1.66667	1.72403
31	0	8	1	1	0	0	1.0	.4	5.55999	5.49630	5.75556	18.19998	5.73845
31	0	1	4	1	0	0	1.0	.4	5.55999	5.49312	5.75556	18.19998	5.74803
15	0	4	4	1	0	0	1.0	.4	4.01850	4.00458	4.33333	9.66667	4.12049
15	0	8	4	1	0	0	1.0	.4	4.01850	4.00980	4.33333	9.66667	4.10018
17	0	8	4	1	0	0	9.0	.4	2.37419	2.34650	3.16790	2.67407	3.07738
31	0	1	8	1	0	0	1.0	.4	5.55999	5.49630	5.75556	18.19998	5.73833
15	0	4	8	1	0	0	1.0	.4	4.01850	4.00980	4.33333	9.66667	4.10018
15	0	8	8	1	0	0	1.0	.4	4.01850	4.01500	4.33333	9.66667	4.07941
15	0	4	4	2	0	0	1.0	.4	8.21505	8.19528	8.33333	33.66664	8.32049
15	0	8	4	2	0	0	1.0	.4	8.21505	8.20267	8.33333	33.66664	8.30017
15	0	4	8	2	0	0	1.0	.4	8.21505	8.20267	8.33333	33.66664	8.30017
15	0	8	8	2	0	0	1.0	.4	8.21505	8.21006	8.33333	33.66664	8.28003
15	0	8	8	2	0	0	9.0	.4	3.10693	3.08396	3.59259	5.22222	3.40697
31	0	4	1	0	1	0	1.0	.4	5.55999	5.49312	5.75556	18.19998	5.74806
31	0	8	1	0	1	0	1.0	.4	5.55999	5.49630	5.75556	18.19998	5.73836
15	0	8	4	2	0	0	9.0	.4	3.21439	3.16252	3.67160	5.69629	3.60627

*L.S. = Least Squares Estimation

try as many parameter sets for the two variable case as for the one and three variable cases. The one variable case was investigated thoroughly because it has a smaller parameter set, and the three variable case is more interesting than the two variable case.

4.6 Comparison of J Values for the Three Variable Case

The same things are compared in the three variable case as in the two variable case. Once again the Taylor Series approximation to our J is used. For all comparisons, we let $\mu_2 = .4$ for our procedure and $\beta_0 = 0$ for all procedures.

We can see that in some cases our procedure and the least squares estimation with an all bias design are very close together. There are parameter sets, however, where there exists more difference between the two methods. Some of these greater differences can be observed at the end of the tables when the first order coefficients become larger relative to the second order coefficients and/or when the variance becomes larger.

TABLE VIIa. Comparison of J Values for Three Variables

$\beta_0 = 0 \quad \mu_2 = .4$

N	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	σ^2	Optimum J for L.S.*	Shrinkage J using $\mu_2 = .4$	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
271	4.0	8.0	1.0	0	0	0	0	0	0	9.0	2.00000	1.00670	4.00000	2.00000	1.90208
271	8.0	8.0	1.0	0	0	0	0	0	0	9.0	2.00000	1.00671	4.00000	2.00000	1.88230
271	4.0	1.0	4.0	0	0	0	0	0	0	9.0	2.00000	1.00669	4.00000	2.00000	1.92198
271	8.0	1.0	4.0	0	0	0	0	0	0	9.0	2.00000	1.00671	4.00000	2.00000	1.90206
27	2.3	1.9	1.6	.5	0	.7	.1	.2	0	2.3	4.30605	4.20056	4.85600	10.53599	4.83703
38	2.3	1.9	1.6	.5	0	.7	.1	.2	0	3.2	4.27965	4.17267	4.83662	10.34280	4.82110
47	2.3	1.9	1.6	.5	0	.7	.1	.2	0	4.0	4.28176	4.17490	4.83816	10.35816	4.82087
31	2.3	1.5	1.8	.5	0	.7	.1	.2	0	2.3	4.47430	4.36832	4.98281	11.80059	4.95792
44	2.3	1.5	1.8	.5	0	.7	.1	.2	0	3.2	4.45596	4.34901	4.96872	11.66008	4.94512
54	2.3	1.5	1.8	.5	0	.7	.1	.2	0	4.0	4.44848	4.34114	4.96300	11.60300	4.93989
31	2.0	1.5	2.0	.5	0	.7	.1	.2	0	2.3	4.47430	4.36794	4.98281	11.80059	4.96003
44	2.0	1.5	2.0	.5	0	.7	.1	.2	0	3.2	4.45596	4.34862	4.96872	11.66008	4.94713
54	2.0	1.5	2.0	.5	0	.7	.1	.2	0	4.0	4.44848	4.34075	4.96300	11.60300	4.94192
31	2.3	1.5	2.0	.5	0	.7	.1	.2	0	2.3	4.47430	4.37480	4.98281	11.80059	4.93579
44	2.3	1.5	2.0	.5	0	.7	.1	.2	0	3.2	4.45596	4.35555	4.96872	11.66008	4.92284
54	2.3	1.5	2.0	.5	0	.7	.1	.2	0	4.0	4.44848	4.34770	4.96300	11.60300	4.91754
31	2.3	1.5	1.6	.4	0	.7	.1	.2	0	2.3	4.31162	4.20344	4.87259	10.28197	4.86271

*L.S. = Least Squares Estimation

TABLE VIIb. Comparison of J Values for Three Variables

$\beta_0 = 0 \quad \mu_2 = .4$

N	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	σ^2	Optimum J for L.S.*	Shrinkage J using $\mu_2 = .4$	*L.S. J for all Bias Design	*L.S. J for all Var Design	J for Proposed Estimator
44	2.3	1.5	1.6	.4	0	.7	.1	.2	0	3.2	4.29464	4.18550	4.86008	10.16324	4.85187
54	2.3	1.5	1.6	.4	0	.7	.1	.2	0	4.0	4.28772	4.17819	4.85500	10.11500	4.84742
27	2.0	1.9	1.6	.4	0	.7	.1	.2	0	2.3	4.15560	4.05255	4.76000	9.21333	4.75591
38	2.0	1.9	1.6	.4	0	.7	.1	.2	0	3.2	4.13106	4.02615	4.74280	9.05007	4.74271
47	2.0	1.9	1.6	.4	0	.7	.1	.2	0	4.0	4.13302	4.02826	4.74416	9.06305	4.74217
27	2.3	1.9	1.6	.4	0	.7	.1	.2	0	2.3	4.15560	4.05948	4.76000	9.21333	4.72877
38	2.3	1.9	1.6	.4	0	.7	.1	.2	0	3.2	4.13106	4.03318	4.74280	9.05007	4.71528
47	2.3	1.9	1.6	.4	0	.7	.1	.2	0	4.0	4.13302	4.03528	4.74416	9.06305	4.71486
31	2.0	1.5	1.8	.4	0	.7	.1	.2	0	2.3	4.31162	4.20534	4.87259	10.28197	4.85785
44	2.0	1.5	1.8	.4	0	.7	.1	.2	0	3.2	4.29464	4.18742	4.86008	10.16324	4.84689
54	2.0	1.5	1.8	.4	0	.7	.1	.2	0	4.0	4.28772	4.18011	4.85500	10.11500	4.84246
31	2.3	1.5	1.8	.4	0	.7	.1	.2	0	2.3	4.31162	4.21182	4.87259	10.28197	4.83362
44	2.3	1.5	1.8	.4	0	.7	.1	.2	0	3.2	4.29464	4.19396	4.86008	10.16324	4.82260
54	2.3	1.5	1.8	.4	0	.7	.1	.2	0	4.0	4.28772	4.18668	4.85500	10.11500	4.81807
31	1.7	1.7	2.0	.4	0	.7	.1	.2	0	2.3	4.31162	4.20149	4.87259	10.28197	4.87140
31	2.0	1.5	2.0	.4	0	.7	.1	.2	0	2.3	4.31162	4.21146	4.87259	10.28197	4.83573
44	2.0	1.5	2.0	.4	0	.7	.1	.2	0	3.2	4.29464	4.19360	4.86008	10.16324	4.82461

*L.S. = Least Squares Estimation

TABLE VIIc. Comparison of J Values for Three Variables

$\beta_0 = 0$ $\mu_2 = .4$

N	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	σ^2	Optimum J for L.S.*	Shrinkage J using $\mu_2 = .4$	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
54	2.0	1.5	2.0	.4	0	.7	.1	.2	0	4.0	4.28772	4.18631	4.85500	10.11500	4.82012
31	2.3	1.5	2.0	.4	0	.7	.1	.2	0	2.3	4.31162	4.21793	4.87259	10.28197	4.81148
44	2.3	1.5	2.0	.4	0	.7	.1	.2	0	3.2	4.29464	4.20012	4.86008	10.16324	4.80032
54	2.3	1.5	2.0	.4	0	.7	.1	.2	0	4.0	4.28772	4.19286	4.85500	10.11500	4.79574
27	2.3	1.9	1.6	.6	0	.7	.1	.2	0	2.3	4.47124	4.35982	4.97333	11.98666	4.96770
38	2.3	1.9	1.6	.6	0	.7	.1	.2	0	3.2	4.44251	4.32947	4.95130	11.76063	4.94881
47	2.3	1.9	1.6	.6	0	.7	.1	.2	0	4.0	4.44480	4.33189	4.95305	11.77860	4.94881
54	2.3	1.5	1.8	.6	0	.7	.1	.2	0	4.0	4.62666	4.51352	5.09500	13.23499	5.08689
31	2.0	1.5	2.0	.6	0	.7	.1	.2	0	2.3	4.65490	4.54282	5.11753	13.46617	5.11006
44	2.0	1.5	2.0	.6	0	.7	.1	.2	0	3.2	4.63484	4.52171	5.10151	13.30178	5.09501
54	2.0	1.5	2.0	.6	0	.7	.1	.2	0	4.0	4.62666	4.51310	5.09500	13.23499	5.08892
31	2.3	1.5	2.0	.6	0	.7	.1	.2	0	2.3	4.65490	4.55005	5.11753	13.46617	5.08581
44	2.3	1.5	2.0	.6	0	.7	.1	.2	0	3.2	4.63484	4.52901	5.10151	13.30178	5.07072
54	2.3	1.5	2.0	.6	0	.7	.1	.2	0	4.0	4.62666	4.52043	5.09500	13.23499	5.06454
44	2.3	1.5	2.0	.4	.6	.4	.1	.2	0	3.2	4.47016	4.35933	4.89629	14.72619	4.88180
54	2.3	1.5	2.0	.4	.6	.4	.1	.2	0	4.0	4.46328	4.35202	4.89100	14.65098	4.87674
27	2.3	1.9	1.6	.5	.4	.4	.1	.2	0	2.3	4.17258	4.06115	4.67466	11.68799	4.06115

*L.S. = Least Squares Estimation

TABLE VIId. Comparison of J Values for Three Variables

$\beta_0 = 0$ $\mu_2 = .4$

N	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	σ^2	Optimum J for L.S.*	Shrinkage J using $\mu_2 = .4$	*L.S. J for all Bias Design	*L.S. J for all Var Design	J for Proposed Estimator
38	2.3	1.9	1.6	.5	.4	.4	.1	.2	0	3.2	4.15060	4.03756	4.65940	11.46872	4.65690
47	2.3	1.9	1.6	.5	.4	.4	.1	.2	0	4.0	4.15236	4.03944	4.66061	11.48616	4.65637
31	2.0	1.5	2.0	.5	.4	.4	.1	.2	0	2.3	4.31199	4.19991	4.77462	13.12325	4.76714
44	2.0	1.5	2.0	.5	.4	.4	.1	.2	0	3.2	4.29684	4.18371	4.76351	12.93780	4.75701
54	2.0	1.5	2.0	.5	.4	.4	.1	.2	0	4.0	4.29066	4.17710	4.75900	12.89899	4.75292
31	2.3	1.5	2.0	.5	.4	.4	.1	.2	0	2.3	4.31199	4.20713	4.77462	13.12325	4.74290
44	2.3	1.5	2.0	.5	.4	.4	.1	.2	0	3.2	4.29684	4.19101	4.76351	12.96378	4.73272
54	2.3	1.5	2.0	.5	.4	.4	.1	.2	0	4.0	4.29066	4.18443	4.75900	12.89899	4.72854
27	2.3	1.9	1.6	.4	.4	.4	.1	.2	0	2.3	4.02872	3.92323	4.57866	10.25866	4.55970
38	2.3	1.9	1.6	.4	.4	.4	.1	.2	0	3.2	4.00860	3.90162	4.56557	10.07174	4.55004
47	2.3	1.9	1.6	.4	.4	.4	.1	.2	0	4.0	4.01021	3.90335	4.56661	10.08661	4.54931
31	2.0	1.5	1.8	.4	.4	.4	.1	.2	0	2.3	4.15588	4.04302	4.66439	11.48216	4.66373
31	2.3	1.5	1.8	.4	.4	.4	.1	.2	0	2.3	4.15588	4.04990	4.66439	11.48216	4.63950
44	2.3	1.5	1.8	.4	.4	.4	.1	.2	0	3.2	4.14210	4.03516	4.65487	11.34622	4.63126
54	2.3	1.5	1.8	.4	.4	.4	.1	.2	0	4.0	4.13648	4.02914	4.65100	11.29099	4.62789

*L.S. = Least Squares Estimation

TABLE VIIe. Comparison of J Values for Three Variables

$\beta_0 = 0 \quad \mu_2 = .4$

N	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	σ^2	Optimum J for L.S.*	Shrinkage J using μ_2 = .4	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
44	2.0	1.5	2.0	.4	.4	.4	.1	.2	0	3.2	4.14210	4.03477	4.65487	11.34622	4.63328
54	2.0	1.5	2.0	.4	.4	.4	.1	.2	0	4.0	4.13648	4.02875	4.65100	11.29099	4.62992
31	2.3	1.5	2.0	.4	.4	.4	.1	.2	0	2.3	4.15588	4.05638	4.66439	11.48216	4.61737
31	2.3	1.5	2.0	.6	.4	.4	.1	.2	0	2.3	4.48703	4.37726	4.90933	14.91129	4.89415
27	2.0	1.9	1.6	.5	0	.4	.1	.2	0	2.3	3.76569	3.67548	4.50400	6.82399	4.47857
31	2.3	1.5	2.0	.5	0	.4	.1	.2	0	2.3	3.88763	3.80709	4.57867	7.53866	4.49306
44	2.3	1.5	2.0	.5	0	.4	.1	.2	0	3.2	3.87445	3.79428	4.57037	7.45926	4.48646
54	2.3	1.5	2.0	.5	0	.4	.1	.2	0	4.0	3.86907	3.79066	4.56700	7.42700	4.48363
27	2.3	1.9	1.6	.4	0	.4	.1	.2	0	2.3	3.58670	3.50997	4.40800	5.82133	4.34636
38	2.3	1.9	1.6	.4	0	.4	.1	.2	0	3.2	3.56927	3.49158	4.39876	5.73484	4.34153
47	2.3	1.9	1.6	.4	0	.4	.1	.2	0	4.0	3.57066	3.49305	4.39950	5.74172	4.34042
27	2.0	1.9	1.6	.4	0	.4	.1	.2	0	2.3	3.58670	3.50451	4.40800	5.82133	4.37351
38	2.0	1.9	1.6	.4	0	.4	.1	.2	0	3.2	3.56927	3.48605	4.39876	5.73484	4.36897
47	2.0	1.9	1.6	.4	0	.4	.1	.2	0	4.0	3.57066	3.48753	4.39950	5.74172	4.36774
31	1.7	1.5	1.8	.4	0	.4	.1	.2	0	2.3	3.69615	3.60714	4.46844	6.38745	4.45447

*L.S. = Least Squares Estimation

TABLE VIIf. Comparison of J Values for Three Variables

$\beta_0 = 0 \quad \mu_2 = .4$

N	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	σ^2	Optimum J for L.S.*	Shrinkage J using $\mu_2 = .4$	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
44	1.7	1.5	1.8	.4	0	.4	.1	.2	0	3.2	3.68435	3.59487	4.46173	6.32455	4.45003
54	1.7	1.5	1.8	.4	0	.4	.1	.2	0	4.0	3.67954	3.58986	4.45900	6.29899	4.44826
31	2.0	1.5	1.8	.4	0	.4	.1	.2	0	2.3	3.69615	3.61520	4.46844	6.38745	4.41879
44	2.0	1.5	1.8	.4	0	.4	.1	.2	0	3.2	3.68435	3.60300	4.46173	6.32455	4.41413
54	2.0	1.5	1.8	.4	0	.4	.1	.2	0	4.0	3.67954	3.59803	4.45900	6.29899	4.41226
31	2.3	1.5	1.8	.4	0	.4	.1	.2	0	2.3	3.69615	3.62041	4.46844	6.38745	4.39456
44	2.3	1.5	1.8	.4	0	.4	.1	.2	0	3.2	3.68435	3.60827	4.46173	6.32455	4.38983
54	2.3	1.5	1.8	.4	0	.4	.1	.2	0	4.0	3.67954	3.60332	4.45900	6.29899	4.38789
21	2.3	1.9	1.8	.4	0	.4	.1	.2	0	2.3	3.40490	3.32531	4.31733	4.97214	4.30995
30	2.3	1.9	1.8	.4	0	.4	.1	.2	0	3.2	3.39946	3.31955	4.31401	4.94856	4.30574
38	2.3	1.9	1.8	.4	0	.4	.1	.2	0	4.0	3.41705	3.33819	4.32300	5.02522	4.30262
31	1.7	1.5	2.0	.4	0	.4	.1	.2	0	2.3	3.69615	3.61209	4.46844	6.38745	4.43235
44	1.7	1.5	2.0	.4	0	.4	.1	.2	0	3.2	3.68435	3.59986	4.46173	6.32455	4.42775
54	1.7	1.5	2.0	.4	0	.4	.1	.2	0	4.0	3.67954	3.59488	4.45900	6.29899	4.42591
31	2.0	1.5	2.0	.4	0	.4	.1	.2	0	2.3	3.69615	3.62013	4.46844	6.38745	4.39667

*L.S. = Least Standard Estimator

TABLE VIIg. Comparison of J Values for Three Variables

$\beta_0 = 0 \quad \mu_2 = .4$

N	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	σ^2	Optimum J for L.S.*	Shrinkage J using $\mu_2 = .4$	*L.S. J for All Bias Design	*L.S. J for All Var Design	J for Proposed Estimator
44	2.0	1.5	2.0	.4	0	.4	.1	.2	0	3.2	3.68435	3.60799	4.46173	6.32455	4.39185
54	2.0	1.5	2.0	.4	0	.4	.1	.2	0	4.0	3.67954	3.60303	4.45900	6.29899	4.38992
31	2.3	1.5	2.0	.4	0	.4	.1	.2	0	2.3	3.69615	3.62534	4.46844	6.38745	4.37242
121	8.0	1.5	2.0	.6	0	.7	.1	.2	0	9	4.62100	4.53610	5.09049	13.18876	4.98293
214	8.0	1.5	2.0	.6	0	.7	.1	.2	0	16	4.61391	4.52873	5.08486	13.13096	4.97766
68	8.0	2.0	2.0	.6	0	.7	.1	.2	0	9	3.96674	3.87858	4.61284	8.28789	4.53793
121	8.0	2.0	2.0	.6	0	.7	.1	.2	0	16	3.96761	3.87951	4.61340	8.29368	4.53696
121	2.0	1.5	5.0	.6	0	.7	.1	.2	0	9	4.62100	4.53299	5.09049	13.18876	4.99443
214	2.0	1.5	5.0	.6	0	.7	.1	.2	0	16	4.61391	4.52561	5.08486	13.13096	4.98895
121	8.0	1.5	5.0	.6	0	.7	.1	.2	0	9	4.62100	4.56174	5.09049	13.18876	4.89424
214	8.0	1.5	5.0	.6	0	.7	.1	.2	0	16	4.61391	4.55445	5.08486	13.13096	4.88868
68	2.0	2.0	5.0	.6	0	.7	.1	.2	0	9	3.96674	3.87413	4.61284	8.28789	4.55773
121	2.0	2.0	5.0	.6	0	.7	.1	.2	0	16	3.96761	3.87506	4.61340	8.29368	4.55679
68	8.0	2.0	5.0	.6	0	.7	.1	.2	0	9	3.96674	3.91442	4.61284	8.28789	4.39289

*L.S. = Least Squares Estimation

TABLE VIIh. Comparison of J Values for Three Variables

$\beta_0 = 0 \quad \mu_2 = .4$

N	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	σ^2	Optimum J for L.S.*	Shrinkage J using $\mu_2 = .4$	*L.S. J for All Bias Design	*L.S. J for all Var Design	J for Proposed Estimator
128	8.0	2.0	5.0	.6	0	.7	.1	.2	0	16	3.96761	3.91532	4.61340	8.29368	4.39250
214	2.0	1.5	2.0	.1	0	.7	.1	.2	0	16	3.88601	3.80929	4.66875	6.47319	4.59939
121	8.0	1.5	2.0	.1	0	.7	.1	.2	0	9	3.89131	3.83216	4.67222	6.49642	4.50191
214	8.0	1.5	2.0	.1	0	.7	.1	.2	0	16	3.88601	3.82670	4.66875	6.47319	4.49913
68	8.0	2.0	2.0	.1	0	.7	.1	.2	0	9	3.38725	3.32887	4.37778	4.52691	4.26761
121	8.0	2.0	2.0	.1	0	.7	.1	.2	0	16	3.38794	3.32959	4.37812	4.52923	4.26639
121	2.0	1.5	5.0	.1	0	.7	.1	.2	0	9	3.89131	3.83331	4.67222	6.49642	4.51341
214	2.0	1.5	5.0	.1	0	.7	.1	.2	0	16	3.88601	3.82454	4.66875	6.47319	4.51042
121	8.0	1.5	5.0	.1	0	.7	.1	.2	0	9	3.89131	3.85000	4.67222	6.49642	4.41322
214	8.0	1.5	5.0	.1	0	.7	.1	.2	0	16	3.88601	3.84458	4.66875	6.47319	4.41015
68	2.0	2.0	5.0	.1	0	.7	.1	.2	0	9	3.38725	3.32590	4.37778	4.52691	4.28741
121	2.0	2.0	5.0	.1	0	.7	.1	.2	0	16	3.38794	3.32661	4.37812	4.52923	4.28622
68	8.0	2.0	5.0	.1	0	.7	.1	.2	0	9	3.38725	3.35303	4.37778	4.52691	4.12257
121	8.0	2.0	5.0	.1	0	.7	.1	.2	0	16	3.38794	3.35373	4.37812	4.52923	4.12193

*L.S. = Least Squares Estimation

CHAPTER V

CONCLUSIONS

We have proposed and discussed a new estimation procedure in a regression or response surface problem. We have shown that for some parameter sets the shrinkage approach produces smaller integrated mean square error averaged over the region of interest than do either an all bias design or an all variance design used with least squares estimation. When our procedure does not result in smaller J values, it is usually close to the optimal value. The only time the proposed procedure does not seem to perform well is when the model we are fitting is heavily biased. It would seem, however, that if we do have an unusual amount of bias present that perhaps we should fit an alternative model.

We might also point out that the shrinkage estimation procedure we propose is not difficult for an experimenter to use. The experimenter needs no more observations for our procedure than he does for least squares. He does not need any additional knowledge about his parameters or experimental results than he needs to employ least squares estimation. The only extra work involved in determining the value of a response at a point in the sample space is in computing $\hat{k}$ which is a ratio of quantities which we need to find anyway. Thus altogether we can say that the mechanics of using the new procedure are not prohibitive to the experimenter.

In order for an experimenter to use our procedure to its fullest advantage, we would make the following recommendation. We recommend that he test for bias before deciding on an estimation procedure. If he detects an appreciable amount of bias and is unable to fit an alternative model, we recommend using least squares estimation with a minimum bias design. If the experimenter detects bias and does fit an alternative model or if he fails to detect bias, we surely recommend our procedure. In this way, he will be assured of having either the smallest J or a close competitor of the best three from among the two least squares procedures and our procedure.

One unexplored area of interest is the relationship of design to our procedure. Because of the complexity of the expression for J, it does not seem to be a trivial consideration. In our numerical work we use several design moments. The most used value is $\mu_2 = .4$. This is used throughout the three variable comparisons of J. The logic for this choice is quite simple. Shrinkage estimation is a variance reducing procedure, so it would seem natural to try to reduce bias by selecting a small design moment. We, however, have no proof that $\mu_2 = .4$ is optimal, and suspect that the optimal second order design moment will fluctuate considerably depending on the values of all the parameters in the problem.

Thus, in summary, we can say that there are some unanswered questions. Also, we can say that while our procedure can be better than existing procedures, it is not always better. We feel, however, that our proposed procedure does have a great deal of merit. Therefore, we have not found one clearly optimal way of estimating a response, but we have added another choice to those already in existence.

BIBLIOGRAPHY

Blight, B.J.N. (1971), Some General Results on Reduced Mean Square Error Estimation. The American Statistician, 25(3): 24-25.

Box, G.E.P., and Draper, N.R. (1959), A Basis for the Selection of a Response Surface Design. Journal of the American Statistical Association, 54: 622-654.

Box, G.E.P. and Wilson, K.B. (1951), On the Experimental Attainment of Optimum Conditions. Journal of the Royal Statistical Society, B13: 1-45.

Cote, R., Manson, A.R., and Hader, R.J. (1973), Minimum Bias Approximation of a General Regression Model. Institute of Statistics Technical Report, North Carolina State University.

David, H.A., and Arens, B.E. (1959), Optimal Spacing in Regression Analysis. Annals of Mathematical Statistics, 30: 1072-1081.

Draper, N.R., and Lawrence, W.E. (1965), Designs Which Minimize Model Inadequacies: Cuboidal Regions of Interest. Biometrika, 52: 111-118.

Hotelling, H. (1941), The Experimental Determination of the Maximum of a Function. Annals of Mathematical Statistics, 12: 20-45.

Karson, M.J. (1970), Design Criteria for Minimum Bias Estimation of Response Surfaces. Journal of the American Statistical Association, 65: 1567-1572.

Karson, M.J., Hader, R.J., and Manson, A.R. (1967), Bias and Variance Criteria for Estimators and Designs for Fitting Polynomial Responses. Institute of Statistics Mimeo Series No. 510, North Carolina State University.

________, Manson, A.R., and Hader, R. J. (1969), Minimum Bias Estimation and Experimental Design for Response Surfaces. Technometrics, 11: 461-475.

Kupper, L. L., and Meydrech, E. F. (1972), A New Approach to Mean Square Error Estimation of Response Surfaces. Institute of Statistics Mimeo Series No. 840, North Carolina State University.

________ (1973), A New Approach to Mean Squared Error Estimation of Response Surfaces. Biometrika, 60: 573-579.

Lancaster, H.O. (1969), The Chi-Squared Distribution. John Wiley & Sons, Inc., New York.

Manson, A.R., Hader, R.J., and Karson, M.J. (1972), Minimum Bias Estimation and Experimental Design Applied to Univariate Polynomial Models. Institute of Statistics Mimeo Series No. 826, North Carolina State University.

Markowitz, E. (1968), Minimum Mean Square Error Estimation of the Standard Deviation of the Normal Distribution. The American Statistician, 22(3): 26.

Myers, R. H. (1971), Response Surface Methodology. Allyn and Bacon, Inc., Boston.

Parish, R.G., Manson, A.R., and Hader, R.J. (1973), Minimum Bias Approximation of Models by Polynomials of Low Order. Institute of Statistics Technical Report, North Carolina State University.

Stein, C. (1964), Inadmissibility of the Usual Estimator for the Variance of a Normal Distribution With Unknown Mean. Annals of the Institute of Statistical Mathematics, 16: 155-160.

Stuart, A. (1969), Reduced Mean Square Error Estimation of σ^p in Normal Samples. The American Statistician, 23(4): 27-28.

Thompson, James R. (1968a). Some Shrinkage Techniques for Estimating the Mean. Journal of the American Statistical Association, 63: 113-122.

Thompson, James R. (1968b). Accuracy Burrowing in the Estimation of the Mean by Shrinkage to an Interval. Journal of the American Statistical Association, 63: 953-963.

The vita has been removed from the scanned document

SHRINKAGE ESTIMATION IN RESPONSE

SURFACE ANALYSIS

by

Linda Catron Malone

(ABSTRACT)

Several attempts have been made to find an estimator of a response which will have a smaller integrated mean square error than existing procedures. In this work another such attempt is made by introducing a shrinkage procedure.

Suppose the true functional relationship between a response η and p independent variables is

$$\eta(x) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{i=1}^{p} \beta_{ii} x_i^2 + \sum_{\substack{i=1 \\ i<j}}^{p-1} \sum_{j=1}^{p} \beta_{ij} x_i x_j .$$

We fit a model

$$\hat{y}(x) = \hat{\beta}_0 + \sum_{i=1}^{p} \hat{k}_i \hat{\beta}_i x_i .$$

We show that the k_i which minimize

$$\sum_{i=1}^{p} E(k_i \hat{\beta}_i - \beta_i)^2$$

are of the form

$$k_i = \frac{\mu_2 N \beta_i^2}{\sigma^2 + \mu_2 N \beta_i^2} .$$

We propose estimating k_i by $\hat{k}_i$ where

$$\hat{k}_i = \frac{\mu_2 N \hat{\beta}_i^2}{\hat{\sigma}^2 + \mu_2 N \hat{\beta}_i^2}$$

and where

$$\underline{\hat{\beta}} = (X_1'X_1)^{-1} X_1'\underline{y}$$

and

$$\hat{\sigma}^2 = (\underline{y}'\underline{y} - \underline{\hat{\beta}}'X'\underline{y})/(N-p)$$

are the usual least squares estimators.

The distribution of $\hat{k}_i$ is derived and the probability that $\hat{k}_i$ is closer to the optimal k_i than a k using upper bounds on the parameters is computed. Also an expression for the integrated mean square error of the proposed procedure is found. Various comparisons among least squares estimation minimum variance and minimum bias designs and optimal least squares and shrinkage estimation are made for the one, two, and three variable cases.