

GOOD RIDGE ESTIMATORS BASED ON PRIOR INFORMATION

by

BENEE F. SWINDEL

Institute of Statistics
Mimeograph Series No. 946
Raleigh, N. C.

GOOD RIDGE ESTIMATORS BASED ON PRIOR INFORMATION

By Binee F. Swindel
U.S. Forest Service and North Carolina State University

SUMMARY

Ridge regression is re-examined and ridge estimators based on prior information are introduced. A necessary and sufficient condition is given for such ridge estimators to yield estimators of every non-null linear combination of the regression coefficients with smaller mean square error than that of the Gauss-Markov best linear unbiased estimator.

Some key words: Biased estimation; Linear models; Mean square error;
Ridge regression.

§1 INTRODUCTION

Unbiased estimators, including

$$\hat{\underline{\beta}}_0 = (X'X)^{-1} X'\underline{y} , \quad (1.1)$$

the best linear unbiased estimator (BLUE), of the slope parameters in the full rank linear model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} , \quad \underline{\varepsilon} \sim (0, I\sigma^2), \quad 0 < \sigma^2 , \quad (1.2)$$

are unreliable if $X_{n \times p}$ is poorly designed (cf. e.g., Stein, 1960; Hoerl and Kennard, 1970a, 1970b; Marquardt 1970, Wonnacott and Wonnacott, 1970; Baranchik 1973; Mayer and Willke, 1973; Swindel, 1974). Consequently, there is considerable interest in various biased estimators of $\underline{\beta}$. The best known of these are the ridge estimators of Hoerl and Kennard. Their family

$$\hat{\underline{\beta}}_k = (X'X + kI)^{-1} X'\underline{y} , \quad 0 \leq k, \quad (1.3)$$

contains the BLUE when $k = 0$, hence the notation of (1.1). For increasing k $\hat{\underline{\beta}}_k$ approaches $\underline{0}$ whatever the design matrix, X , and the data, \underline{y} .

The very description of the entire family (1.3) suggests that $\hat{\beta}_k$, for some $k \in (0, +\infty)$, might be an effective compromise between $\hat{\beta}_0$ which is unbiased, but unstable (if X is poorly designed) and $\hat{\beta}_{+\infty} \equiv \underline{0}$ which is certainly a stable, but biased estimator of $\underline{\beta}$. Other intuitive justifications for shrinking $\hat{\beta}_0$ have been suggested. Hoerl and Kennard (1970a) showed that the expected length of $\hat{\beta}_0$ exceeds the length of $\underline{\beta}$. They (*loc. cit.*) showed that $\hat{\beta}_k$ is the shortest estimator in an equivalence class of estimators with the same residual sum of squares. And they showed that $\hat{\beta}_k$ has minimum residual sum of squares in an equivalence class of estimators of the same length. None of these arguments are compelling reasons for choosing $\hat{\beta}_k$ with $k > 0$ as an estimator of $\underline{\beta}$. Still, a dynamic reiteration of these results is useful. As k increases without bound from zero, $\hat{\beta}_k$ traces a curved path through the parameter space from $\hat{\beta}_0$ to $\underline{0}$. The path is uniquely determined so that the distance from $\hat{\beta}_k$ to $\underline{0}$ is diminished as rapidly as possible while the residual sum of squares is increased as slowly as possible.

Hoerl and Kennard (1970a) also proved that the expected distance between $\hat{\beta}_k$ and $\underline{\beta}$ must initially decrease as k increases from zero, *i.e.*, there exist always positive values of k such that

$$E(\hat{\beta}_k - \underline{\beta})'(\hat{\beta}_k - \underline{\beta}) < E(\hat{\beta}_0 - \underline{\beta})'(\hat{\beta}_0 - \underline{\beta}). \quad (1.4)$$

Swindel and Chapman (1973) showed explicitly a more stringent property, *viz.*,

$$E(\underline{L}'\hat{\beta}_k - \underline{L}'\underline{\beta})^2 < E(\underline{L}'\hat{\beta}_0 - \underline{L}'\underline{\beta})^2, \quad \forall \underline{L} \neq \underline{0} \quad (1.5)$$

if and only if k is in the never empty open interval

$$K = K(X, \underline{\beta}, \sigma^2) = \begin{cases} (0, +\infty), & \text{if } [(X'X)^{-1} - \underline{\beta}\underline{\beta}'/\sigma^2] \text{ is non-negative} \\ & \text{definite;} \\ \left(0, \frac{-2}{\min \text{ root } [(X'X)^{-1} - \underline{\beta}\underline{\beta}'/\sigma^2]} \right), & \text{otherwise.} \end{cases} \quad (1.6)$$

Thus, $\hat{\underline{\beta}}_k$ provides strictly smaller mean square error estimators of every non-null linear combination of the parameters than does $\hat{\underline{\beta}}_0$ if and only if k is in K . Equivalently, every coordinate of $\hat{\underline{\beta}}_k$ is expected to be strictly closer to the corresponding coordinate of $\underline{\beta}$ than is the corresponding coordinate of $\hat{\underline{\beta}}_0$ if and only if k is in K . These are forceful reasons for examining the family (1.3) for estimators of $\underline{\beta}$ which are not BLUE.

Finally, Hoerl and Kennard (1970b) showed that something apparently can be done even though an optimum choice of k , and hence of $\hat{\underline{\beta}}_k$, depends on the unknown parameters of the model. They believe that a judicious estimator, $\hat{\underline{\beta}}_k$, of $\underline{\beta}$ can be chosen from an inspection of the ridge trace, i.e., the graph of the p coordinates of $\hat{\underline{\beta}}_k$ and the residual sum of squares, $(\underline{y} - X\hat{\underline{\beta}}_k)'(\underline{y} - X\hat{\underline{\beta}}_k)$, for small values of k . They (loc. cit.) presented data for which the ridge trace seemed to stabilize for small values of k and they believe that the corresponding ridge estimator, $\hat{\underline{\beta}}_k$, is a reasonable choice. Similar results have apparently been achieved in some other applications (cf. Talwar and Ashlock, 1970; Rosen and Provder, 1971; Balaban and Haspert, 1972; McDonald and Schwing, 1973). Incidentally, equation (1.6) immediately suggests an interval over which the ridge trace might reasonably be examined, viz., $\hat{K} = K(X, \hat{\underline{\beta}}, \hat{\sigma}^2)$ where $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are estimates of $\underline{\beta}$ and σ^2 , respectively (cf. Swindel and Chapman, 1973).

§2. ALTERATION AND RECONSIDERATION

In the remainder of this paper the foregoing ideas are subjected to a statistically significant alteration and re-examination. An arbitrary point, \underline{b} , in the parameter space is assigned to the role previously assumed by the origin. The resulting family of estimators

$$\underline{\hat{\beta}}_k = (X'X + kI)^{-1} (X'\underline{y} + k\underline{b}), \quad 0 \leq k, \quad (2.1)$$

contains the estimators of Hoerl and Kennard (when $\underline{b} = \underline{0}$) and the Gauss-Markov BLUE (when $k = 0$).

In §3 it will be shown that: $\underline{\hat{\beta}}_k$ approaches \underline{b} strictly monotonically as $k \rightarrow +\infty$, whatever X and \underline{y} ; $\underline{\hat{\beta}}_k$ is closest to \underline{b} in an equivalence class of estimators of $\underline{\beta}$ with the same residual sum of squares; and $\underline{\hat{\beta}}_k$ has minimum residual sum of squares in an equivalence class of estimators of $\underline{\beta}$ which are equal distance from \underline{b} . Thus, as k increases without bound from zero $\underline{\hat{\beta}}_k$ traces a curved path through the parameter space from $\hat{\beta}_0$ to \underline{b} . The path is uniquely determined so that the distance from $\underline{\hat{\beta}}_k$ to \underline{b} is diminished as rapidly as possible while the residual sum of squares is increased as slowly as possible.

With $\underline{\hat{\beta}}_k$ so inclined to \underline{b} , in the applications \underline{b} might well be chosen to reflect as well as possible the prior information or hypotheses on $\underline{\beta}$. Formally then, call \underline{b} the prior information on $\underline{\beta}$ and $\underline{\hat{\beta}}_k$, $0 \leq k$, the family of ridge estimators of $\underline{\beta}$ based on prior information, \underline{b} . Given this determination of \underline{b} , the choice of a particular k , and hence $\underline{\hat{\beta}}_k$, is seen to be a choice of compromise between $\hat{\beta}_0$ (the BLUE of $\underline{\beta}$ based on the sample data) and $\underline{\hat{\beta}}_{+\infty} \equiv \underline{b}$ (the prior

information). And the approach of $\hat{\underline{b}}_k$ to \underline{b} requires no pleading, i.e., the intuitive arguments for shrinking $\hat{\underline{\beta}}_0$ toward \underline{b} are substantially stronger than those for shrinking $\hat{\underline{\beta}}_0$ toward $\underline{0}$ ^{1/}.

In §4 generalizations of (1.5) and (1.6) are given. These are the main results of this paper. They are briefly discussed in §5.

In §6, a still more general formulation is mentioned. There \underline{b} is considered to be the realized value of a p-dimensional random variable, \underline{B} . Some statistics for this case are recorded along with some generalizations of preceding results. It will easily be seen that the ridge estimator based on prior information, \underline{B} , is a biased estimator of $\underline{\beta}$ only if \underline{B} is biased.

§3. THE FAMILY $\hat{\underline{b}}_k$ AS A TRAJECTORY IN PARAMETER SPACE

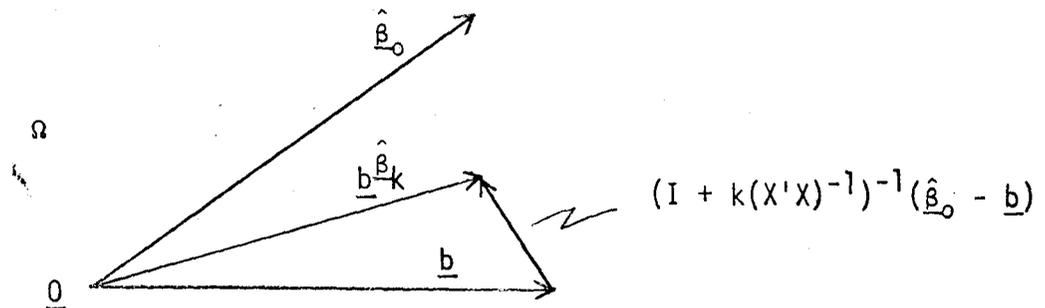
The parameter space for the slope parameters in (1.2) is $\Omega = p$ -dimensional Euclidean space. The origin, $\underline{0}$, is fixed, and \underline{b} and $\underline{\beta}$ (unknown) are other fixed points. $\hat{\underline{b}}_k$ defined by (2.1) is a p-dimensional random

^{1/}I. J. Good suggested, in a personal communication, that fixing $\underline{b} = \underline{0}$ is, in a way, invoking the least informative prior information (in the sense that $\underline{0}$ represents prior information least informative physically about the effects on \underline{y} produced by the variables in X), i.e., it is the classical null hypothesis.

variable related to \underline{b} (the prior information) and to $\hat{\underline{\beta}}_0 - \underline{b}$ (the difference between the BLUE and the prior information) as in Figure 1, i.e.,

$$\begin{aligned}
 \underline{b}_{\hat{\underline{\beta}}_k} &= (X'X + kI)^{-1} (X'y + kb) \\
 &= (X'X + kI)^{-1} X'X(X'X)^{-1} (X'y + kb) \\
 &= (I + k(X'X)^{-1})^{-1} (\hat{\underline{\beta}}_0 - \underline{b} + \underline{b} + k(X'X)^{-1}\underline{b}) \\
 &= (I + k(X'X)^{-1})^{-1} (\hat{\underline{\beta}}_0 - \underline{b}) + \underline{b}. \tag{3.1}
 \end{aligned}$$

Figure 1



Clearly, $\underline{b}_{\hat{\underline{\beta}}_k} \equiv \underline{b}$ if $\hat{\underline{\beta}}_0 = \underline{b}$, i.e., if the BLUE and the prior information agree, then $\underline{b}_{\hat{\underline{\beta}}_k}$ agrees with them both for all k .

So, hereafter, assume $\hat{\underline{\beta}}_0 \neq \underline{b}$. Then

$$\begin{aligned}
 \text{i) } \underline{b}_{\hat{\underline{\beta}}_0} &= \hat{\underline{\beta}}_0 \quad \text{by (3.1).} \\
 \text{ii) } \underline{b}_{\hat{\underline{\beta}}_{+\infty}} &= \lim_{k \rightarrow +\infty} \underline{b}_{\hat{\underline{\beta}}_k} \\
 &= \underline{b} + \lim_{k \rightarrow +\infty} (I + k(X'X)^{-1})^{-1} (\hat{\underline{\beta}}_0 - \underline{b}) \\
 &= \underline{b}
 \end{aligned}$$

by the fact that $(I + k(X'X)^{-1})^{-1}$ is a positive definite matrix for all non-negative k with roots all converging to zero as k increases without bound (Hoerl and Kennard, 1970a), i.e., $(I + k(X'X)^{-1})^{-1}$ converges to a null matrix as $k \rightarrow +\infty$.

iii) The Euclidean distance, $\| \underline{\hat{\beta}}_k - \underline{b} \|^2$, between $\underline{\hat{\beta}}_k$ and \underline{b} , i.e., $\| (I + k(X'X)^{-1})^{-1} (\underline{\hat{\beta}}_0 - \underline{b}) \|^2$, is strictly monotonically decreasing as k increases. A proof using only the facts that the sum (Bellman, 1970, page 117) and the product (Roy, 1954) of two symmetric, positive definite matrices is positive definite is

$$\begin{aligned}
 & 0 \leq k \quad \text{and} \quad 0 < \Delta k \\
 \Rightarrow & \Delta k (X'X)^{-1} \text{ is positive definite (p.d.)} \\
 \Rightarrow & I + [I + k(X'X)^{-1}]^{-1} \Delta k (X'X)^{-1} - I \text{ is p.d.} \\
 \Rightarrow & [I + k(X'X)^{-1}]^{-1} [I + (k + \Delta k)(X'X)^{-1}] - I \text{ is p.d.} \\
 \Rightarrow & [I + k(X'X)^{-1}]^{-1} - [I + (k + \Delta k)(X'X)^{-1}]^{-1} \text{ is p.d.} \\
 \Rightarrow & \{ [I + k(X'X)^{-1}]^{-1} + [I + (k + \Delta k)(X'X)^{-1}]^{-1} \} \text{ times} \\
 & \quad \{ [I + k(X'X)^{-1}]^{-1} - [I + (k + \Delta k)(X'X)^{-1}]^{-1} \} \text{ is p.d.} \\
 \Rightarrow & [I + k(X'X)^{-1}]^{-2} - [I + (k + \Delta k)(X'X)^{-1}]^{-2} \text{ is p.d.} \\
 \Rightarrow & (\underline{\hat{\beta}}_0 - \underline{b})' [I + k(X'X)^{-1}]^{-2} (\underline{\hat{\beta}}_0 - \underline{b}) \\
 & \quad - (\underline{\hat{\beta}}_0 - \underline{b})' [I + (k + \Delta k)(X'X)^{-1}]^{-2} (\underline{\hat{\beta}}_0 - \underline{b}) > 0,
 \end{aligned}$$

i.e., $0 \leq k$ and $0 \leq \Delta k$ implies

$$\begin{aligned}
 & \| [I + k(X'X)^{-1}]^{-1} (\underline{\hat{\beta}}_0 - \underline{b}) \|^2 \\
 & > \| [I + (k + \Delta k)(X'X)^{-1}]^{-1} (\underline{\hat{\beta}}_0 - \underline{b}) \|^2.
 \end{aligned}$$

iv) $\underline{b}_{\hat{\beta}_k}$ minimizes $(\hat{\beta} - \underline{b})'(\hat{\beta} - \underline{b})$ such that $(\underline{y} - X\hat{\beta})'(\underline{y} - X\hat{\beta}) = c$.

Write the objective function

$$\psi = (\hat{\beta} - \underline{b})'(\hat{\beta} - \underline{b}) + \frac{1}{k} \left[(\underline{y} - X\hat{\beta})'(\underline{y} - X\hat{\beta}) - c \right]$$

where $1/k$ is a Lagrangian multiplier. Observe that ψ is a strictly convex function of $\hat{\beta}$ (cf. Kunzi, Krelle, and Oettli, 1966, page 39). Differentiating ψ with respect to $\hat{\beta}$ and equating the results to $\underline{0}$ verifies the claim.

v) $\underline{b}_{\hat{\beta}_k}$ minimizes $(\underline{y} - X\hat{\beta})'(\underline{y} - X\hat{\beta})$ such that $(\hat{\beta} - \underline{b})'(\hat{\beta} - \underline{b}) = c$. Verification proceeds as in iv) with the objective function now taken to be

$$\psi = (\underline{y} - X\hat{\beta})'(\underline{y} - X\hat{\beta}) + k \left[(\hat{\beta} - \underline{b})'(\hat{\beta} - \underline{b}) - c \right].$$

§4. A NECESSARY AND SUFFICIENT CONDITION FOR $\underline{b}_{\hat{\beta}_k}$

TO BE A GOOD RIDGE ESTIMATOR OF $\underline{\beta}$ BASED ON PRIOR INFORMATION, \underline{b}

Following Swindel and Chapman (1973), call $\underline{b}_{\hat{\beta}_k}$ a good ridge estimator of $\underline{\beta}$ based on prior information, \underline{b} , if it provides a strictly smaller mean square error estimator of every non-null linear combination of the parameters, $\underline{\beta}$, than does the Gauss-Markov estimator, $\hat{\beta}_0$, i.e., if

$$E(\underline{L}'\underline{b}_{\hat{\beta}_k} - \underline{L}'\underline{\beta})^2 < E(\underline{L}'\hat{\beta}_0 - \underline{L}'\underline{\beta}), \quad \forall \underline{L} \neq \underline{0}. \quad (4.1)$$

A necessary and sufficient condition for (4.1) is that k is in the open interval

$$K = K(\underline{b}, X, \underline{\beta}, \sigma^2) = \begin{cases} (0, +\infty), & \text{if } \left[(X'X)^{-1} - (\underline{b} - \underline{\beta})(\underline{b} - \underline{\beta})'/\sigma^2 \right] \\ & \text{is non-negative definite;} \\ 0, & \frac{-2}{\min \text{ root } \left[(X'X)^{-1} - (\underline{b} - \underline{\beta})(\underline{b} - \underline{\beta})'/\sigma^2 \right]} \end{cases}, \text{ otherwise.} \quad (4.2)$$

A proof is given in §6.

§5. COMMENTS

If the prior information, \underline{b} , is sufficiently close to $\underline{\beta}$, then every compromise via (2.1) of the BLUE and the prior information is preferable to the BLUE with regard to small mean square error estimation. In particular, it is easy to show that

$$(\underline{b} - \underline{\beta})'(\underline{b} - \underline{\beta}) \leq \sigma^2 \min \text{root } (X'X)^{-1} \quad (5.1)$$

is a sufficient condition for $(X'X)^{-1} - (\underline{b} - \underline{\beta})(\underline{b} - \underline{\beta})'/\sigma^2$ to be non-negative definite and hence for every $\hat{\underline{\beta}}_k$, $0 \neq k$, to be a good estimator according to (4.1).

$K(\underline{b}, X, \underline{\beta}, \sigma^2)$ is never empty whatever the prior information, the design matrix, and the parameters of the model, *i.e.*, there always exist compromises via (2.1) of the BLUE with the prior information which are preferable to the BLUE if the criterion is small mean square error estimation. To paraphrase Efron and Morris (1973), even if ($p = 1$ and) $\underline{\beta}$ is the tea consumption in Taiwan or the weight of hogs in Montana and \underline{b} is the speed of light, it is still possible to incline $\hat{\underline{\beta}}_0$ a little toward \underline{b} and thereby improve the mean square error of estimation. In this respect §4 is somewhat reminiscent of the surprising results of James and Stein (1961).

§6. PRIOR INFORMATION AS A RANDOM VARIABLE

It seems reasonable in the applications, when confronted with the analysis of data satisfying (1.2), to consider the prior information, \underline{b} , on $\underline{\beta}$ as fixed

and to examine estimators of $\underline{\beta}$ conditioned on \underline{b} . So, the level of generality of §4 seems to me to be the most useful. For the sake of proofs, however, it's not hard to generalize that formulation by taking \underline{b} to be the realized value of a p -dimensional random variable, \underline{B} .

$$\underline{B}\hat{\underline{\beta}}_k = (X'X + kI)^{-1} (X'\underline{y} + k\underline{B}), \quad 0 \leq k, \quad (6.1)$$

is the family of ridge estimators of $\underline{\beta}$ based on data \underline{y} and prior information, \underline{B} , and $\underline{L}'\underline{B}\hat{\underline{\beta}}_k$, $0 \leq k$, is the corresponding family of estimators of $\underline{L}'\underline{\beta}$ where $\underline{L} \neq \underline{0}$ is arbitrary.

Define

$$W_k = (X'X + kI)^{-1}, \quad Z_k = [I + k(X'X)^{-1}]^{-1} \quad (6.2)$$

and notice (Hoerl and Kennard (1970a)) that

$$Z_k = W_k X'X = I - kW_k. \quad (6.3)$$

Assume \underline{B} has finite expectation. Then

$$\underline{L}'\underline{B}\hat{\underline{\beta}}_k = \underline{L}'W_k X'\underline{y} + k\underline{L}'W_k \underline{B} \quad \text{and} \quad (6.4)$$

$$E(\underline{L}'\underline{B}\hat{\underline{\beta}}_k) = \underline{L}'Z_k \underline{\beta} + k\underline{L}'W_k E(\underline{B}) \quad (6.5)$$

by (6.2), (6.1), and (1.2). Thus, the bias of $\underline{L}'\underline{B}\hat{\underline{\beta}}_k$ is

$$\begin{aligned} B(\underline{L}'\underline{B}\hat{\underline{\beta}}_k) &= E(\underline{L}'\underline{B}\hat{\underline{\beta}}_k) - \underline{L}'\underline{\beta} \\ &= \underline{L}'(Z_k - I)\underline{\beta} + k\underline{L}'W_k E(\underline{B}) \\ &= k \underline{L}' W_k [E(\underline{B}) - \underline{\beta}] \end{aligned} \quad (6.6)$$

by (6.3) and $\underline{L}'\underline{B}\hat{\beta}_k$ is unbiased if \underline{B} is.

Assume, moreover, that \underline{B} is independent of \underline{y} and has a finite covariance matrix, $V(\underline{B})$. Then the variance of $\underline{L}'\underline{B}\hat{\beta}_k$ is

$$\begin{aligned} V(\underline{L}'\underline{B}\hat{\beta}_k) &= \sigma^2 \underline{L}'W_k X'XW_k \underline{L} + k^2 \underline{L}'W_k V(\underline{B})W_k \underline{L} \\ &= \underline{L}'W_k \left[\sigma^2 X'X + k^2 V(\underline{B}) \right] W_k \underline{L} \end{aligned} \quad (6.7)$$

by (6.4) and (1.2). The mean square error of $\underline{L}'\underline{B}\hat{\beta}_k$ is

$$\begin{aligned} S(\underline{L}'\underline{B}\hat{\beta}_k) &= V(\underline{L}'\underline{B}\hat{\beta}_k) + B^2(\underline{L}'\underline{B}\hat{\beta}_k) \\ &= \underline{L}'W_k \left\{ \sigma^2 X'X + k^2 V(\underline{B}) + k^2 [E(\underline{B}) - \underline{\beta}][E(\underline{B}) - \underline{\beta}]' \right\} W_k \underline{L} \end{aligned}$$

by (6.7) and (6.6). Denoting

$$S(\underline{B}) = V(\underline{B}) + [E(\underline{B}) - \underline{\beta}][E(\underline{B}) - \underline{\beta}]' \quad (6.8)$$

we have

$$S(\underline{L}'\underline{B}\hat{\beta}_k) = \underline{L}'W_k \left[\sigma^2 X'X + k^2 S(\underline{B}) \right] W_k \underline{L}. \quad (6.9)$$

The mean square error of the BLUE of $\underline{L}'\underline{\beta}$ is, of course,

$$S(\underline{L}'\hat{\beta}_0) = \sigma^2 \underline{L}'(X'X)^{-1} \underline{L}. \quad (6.10)$$

Define $\underline{B}\hat{\beta}_k$ to be a good ridge estimator of $\underline{\beta}$ based on prior information, \underline{B} , if

$$S(\underline{L}'\underline{B}\hat{\beta}_k) < S(\underline{L}'\hat{\beta}_0) \quad \forall \underline{L} \neq \underline{0}. \quad (6.11)$$

A necessary and sufficient condition for (6.11) is (by (6.9) and (6.10))

$$W_k^{-1}(X'X)^{-1}W_k^{-1} - X'X - \frac{k^2}{\sigma^2} S(\underline{B}) \text{ is p.d.}$$

or, equivalently (by (6.2)),

$$2I + k[(X'X)^{-1} - S(\underline{B})/\sigma^2] \text{ is p.d.}$$

or, equivalently,

$$\begin{aligned} 0 < \min \text{root} \left\{ 2I + k[(X'X)^{-1} - S(\underline{B})/\sigma^2] \right\} \\ = 2 + k \min \text{root} \left[(X'X)^{-1} - S(\underline{B})/\sigma^2 \right]. \end{aligned} \quad (6.12)$$

Clearly, a necessary and sufficient condition for (6.12) and hence for (6.11) is that k is in

$$K = K(S(\underline{B}), X, \sigma^2) = \begin{cases} (0, +\infty), & \text{if } [(X'X)^{-1} - S(\underline{B})/\sigma^2] \\ & \text{is non-negative definite;} \\ \left(0, \frac{-2}{\min \text{root} [(X'X)^{-1} - S(\underline{B})/\sigma^2]} \right), & \text{otherwise.} \end{cases} \quad (6.13)$$

Taking \underline{B} to have a distribution degenerate at \underline{b} so that (cf. (6.8))
 $S(\underline{B}) = [\underline{b} - \underline{\beta}][\underline{b} - \underline{\beta}]'$ gives the results claimed in §4.

REFERENCES

- Balaban, H.S. and Haspert, K. (1972) Estimating Weibull parameters for a general class of devices from limited failure data. IEEE Transactions on Reliability R-21, 111-117.
- Baranchik, A.J. (1973) Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables. The Annals of Statistics 1, 312-321.
- Bellman, R. (1970) Introduction to Matrix Analysis, 2nd edition. New York: McGraw-Hill.
- Efron, B. and Morris, C. (1973) Combining possibly related estimation problems. Journal of the Royal Statistical Society B35, 379-421.
- Hoerl, A.E. and Kennard, R.W. (1970a) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12, 55-67.
- Hoerl, A.E. and Kennard, R.W. (1970b) Ridge regression: applications to non-orthogonal problems. Technometrics 12, 69-82.
- James, W. and Stein, C. (1961) Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium 1, 361-379.
- Kunzi, H.P., Krelle, W., and Oettli, W. (1966) Nonlinear programming. Waltham, Mass: Blaisdell.
- Marquardt, D.W. (1970) Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12, 591-612.
- Mayer, L.S. and Willke, T.A. (1973) On biased estimation in linear models. Technometrics 15, 497-508.

- McDonald, G.C. and Schwing, R.C. (1973) Instabilities of regression estimates relating air pollution to mortality. Technometrics 15, 463-481.
- Rosen, E.M. and Provder, T. (1971) The instrument spreading correction in GPC. III. The general shape function using singular value decomposition with a nonlinear calibration curve. Journal of Applied Polymer Science 15, 1687-1702.
- Roy, S.N. (1954) A useful theorem in matrix theory. American Mathematical Society Proceedings 5, 635-638.
- Stein, C. (1960) Multiple Regression. Contributions to Probability and Statistics. Stanford University Press.
- Swindel, B.F. (1974) Instability of regression coefficients illustrated. The American Statistician 28, 63-65.
- Swindel, B.F. and Chapman, D.D. (1973) Good ridge estimators. A paper presented at the annual meeting of the Institute of Mathematical Statistics, the Biometric Society, and the American Statistical Association.
- Talwar, B.L. and Ashlock, L.T. (1970) Selecting moulding conditions for themosets. SPE Journal 26, 42-46.
- Wonnacott, R.J. and Wonnacott, T.H. (1970) Econometrics New York: Wiley and Sons.