The Use of Multiple Measurements to Monitor Protocol Adherence in Epidemiological Studies*

H. M. Schey and C. E. Davis
Department of Biostatistics,
University of North Carolina
Chapel Hill, North Carolina

*This work was supported by the U.S. National Heart, Lung, and Blood Institute Contracts NIH-NHLBI-71-2245 and 5-732-H607005-03 from National Institutes of Health.

1. INTRODUCTION

The problem of estimating the parameters of a regression equation has received considerable attention in the statistical literature. Recently, this interest has focused on biased estimation with the hope that, by abandoning unbiasedness, significant reductions in mean squared error (or mean squared risk) could be achieved. Many ad hoc estimators have been proposed for this purpose including Hoerl and Kennard's (1970) Ridge regression, Marquardt's (1970) Generalized Inverse regression, Goldstein and Smith's (1974) Shrinkage regression, Mayer and Willke's (1973) Shrunken estimators and Toro and Wallace's (1968) False Restrictions regression. This list is by no means exhaustive.

Several methods for generalizing the concept of mean squared error to the multi-parameter case are in common usage. In this work the mean squared error matrix shall be used intact. Optimality of the estimator will be judged with respect to an ordering which will be defined among mean squared error matrices. It is well known that attempts to minimize mean squared error directly invariably results in an estimator which is a function of the parameter being estimated. One common method for overcoming this drawback is to employ the minimax principle. In a modified form, this is the approach taken in this work. A minimax mean squared error estimator is obtained which is based on the entire mean squared error matrix.

In a recent work by Gerig and Zarate (1976), this same problem was addressed using the trace of the mean squared error matrix as a criterion. The resulting Directionally Minimax Trace Mean Squared Error estimator turned out to be from the class of Shrunken regression estimators proposed by Mayer and Willke (1973). Some parallel results in this area are presented by Bunke (1975) and in a survey paper by Rao (1976).

In Section 2, an ordering among n.n.d. matrices is defined and discussed. In Section 3 the concept of a supremum of a set of n.n.d. matrices is introduced and a theorem is given which lists sufficient conditions for a matrix to be supremum. In Section 4 the results of Sections 2 and 3 are used to obtain the Directionally Minimax Mean Squared Error (DMMSE) criterion and from this the DMMSE estimate is ultimately obtained. The resulting estimators are shown to correspond to the Ridge regression estimator of Hoerl and Kennard (1970).

2. ORDERING MEAN SQUARED ERROR MATRICES

In this work competing estimators of the regression parameters will be compared by relating their respective mean squared error matrices. For this purpose it is necessary to define a (partial) ordering of square matrices. If A is a $(p \times p)$ matrix, then we shall say that A is non-negative definite (n.n.d.) and write $A \geq 0$, if $x'Ax \geq 0$ for all $(p \times 1)$ vectors x. We shall define $A \geq B$ (or $B \leq A$) to mean that $A - B \geq 0$. Note that in some instances two matrices may not be comparable with respect to the relation. It can easily be seen that the relation $\geq$ between square matrices is transitive and anti-symmetric.

In view of this a set of square matrices, $\mathcal{G}$, may be said to possess a minimum if there exists a matrix $M \in \mathcal{G}$ such that $A \in \mathcal{G}$ implies $A \geq M$. The anti-symmetry of the relation $\geq$ insures that the minimum is unique.

Geometrically, we can associate a point set, $E_A = \{x | x'Ax = 1\}$, with each square matrix, A. Some intuition can be gained concerning the ordering among n.n.d. matrices from the following theorem.

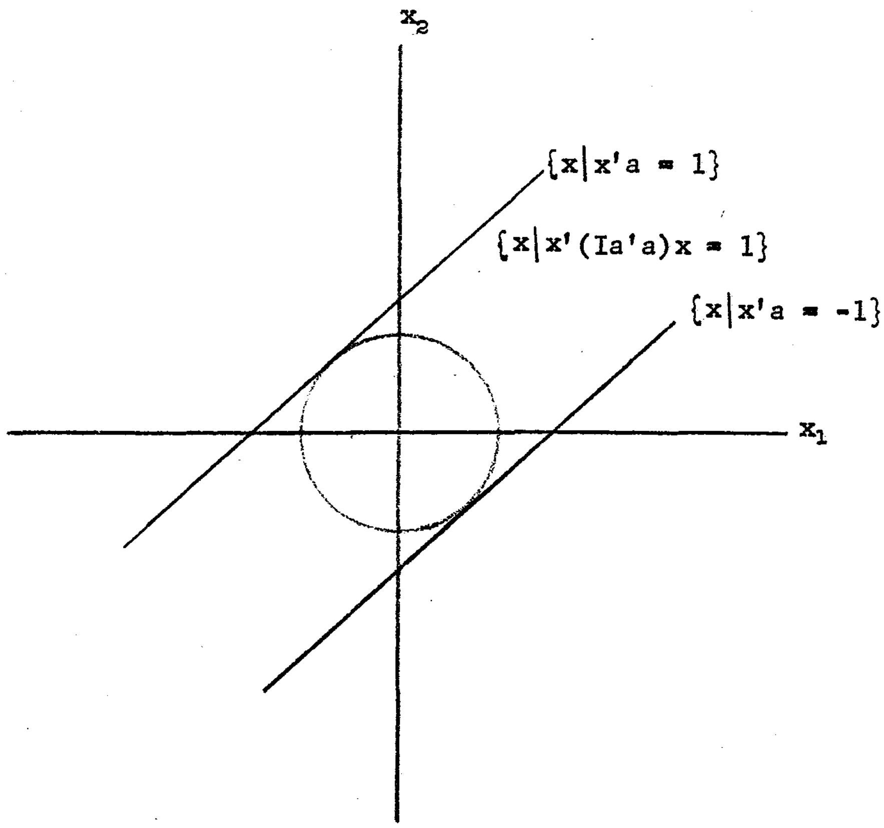
Theorem 2.1: Suppose that A and B are any two n.n.d. matrices, then $x \in E_A$ implies that there exists a number m, $0 < m \leq 1$, such that $m\, x \in E_B$, if, and only if, $A \leq B$.

Proof: Assume first that $A \leq B$ and suppose that $x \in E_A$, then $1 = x'Ax \leq x'Bx$. Since $x'Bx > 0$ we choose $m = (x'Bx)^{-1/2} \leq 1$ and find that $m\, x'Bx\, m = (x'Bx)^{-1}\, x'Bx = 1$, that is, $m\, x \in E_B$, where $0 < m \leq 1$.

Assume now that x is an arbitrary vector. Then either $x'A = 0$ or $m_1 x \in E_A$ for some $m_1 \neq 0$. If the former holds, then $x'Ax = 0 \leq x'Bx$. If the latter holds, then $m_1^2 x'Ax = 1$ and, by hypothesis, there exists $0 < m_2 \leq 1$ such that $m_2 m_1 x'Bx m_1 m_2 = m_1^2 m_2^2 x'Bx = 1$. Now, $m_1^2 x'Ax = 1 = m_1^2 m_2^2 x'Bx$ and, hence, $x'Ax = m_2^2 x'Bx \leq x'Bx$. But since x was chosen arbitrarily $A \leq B$.

By Theorem 2.1 we can conclude intuitively that, for n.n.d. matrices, $A \le B$ means that the point set E_B is entirely contained in E_A, in the sense that, if we choose a point in E_A and travel towards the origin, we will cross E_B before reaching the origin. The following example illustrates the idea in two dimensions.

<u>Example 2.1</u>: Let a be a (2×1) vector, $A = aa'$, and $B = a'aI_2$. It follows from the Cauchy-Schwarz inequality that $x'Ax = x'aa'x \le a'ax'x = x'Bx$, and, hence, that $A \le B$. Note that $E_A = \{x \mid x'a = \pm 1\}$ and E_B is the circle of radius 1 about the origin. Sets E_A and E_B are graphed in the figure. For these specific matrices, the condition of the theorem can be seen to hold.


x2
{x|x'a = 1}
{x|x'(Ia'a)x = 1}
{x|x'a = -1}
x1

3. THE SUPREMUM OF A SET OF MATRICES

It will be necessary in later sections to obtain a matrix which can be interpreted to be the supremum of a set of matrices. The following defines what is meant by a supremum in this context.

<u>Definition 3.1</u>: Let $\mathfrak{V}$ be the set of n.n.d. matrices of a given dimension. A matrix $M \in \mathfrak{V}$ is the supremum of a set $\mathfrak{W} \subseteq \mathfrak{V}$ (with respect to the defined ordering) if, and only if,

(i) $A \in \mathfrak{W}$ implies $A \leq M$, that is M is an upper bound for $\mathfrak{W}$,

and

(ii) if $M_1 \in \mathfrak{V}$, and $A \in \mathfrak{W}$ implies $M_1 \geq A$, then $M_1 \geq M$, that is M is the minimum of the set of upper bounds for $\mathfrak{W}$.

It should be noted that, even when the set $\mathfrak{W}$ is, in some sense, bounded, a supremum may not exist.

Suppose that a n.n.d. matrix, A, is a function of a vector, t, and is denoted by A(t). Define the set $\mathfrak{W} = \{A(t), t \in \mathfrak{T}\}$, where $\mathfrak{T}$ is some non-empty set of vectors t. If a supremum, M, of $\mathfrak{W}$ exists, we shall denote it by

$$M = \sup_{t \in \mathfrak{T}} A(t).$$

The following theorem gives a sufficient condition for a n.n.d. matrix, M, to be the supremum of a set $\mathfrak{W} = \{A(t), t \in \mathfrak{T}\}$ under the assumption that A(t) is n.n.d. for all $t \in \mathfrak{T}$.

<u>Theorem 3.1</u>: Let M and A(t) for all $t \in \mathfrak{J}$ be n.n.d. and assume that

(i) $A(t) \leq M$ for all $t \in \mathfrak{J}$

and

(ii) for every $x \in E_M$ there exists a $t \in \mathfrak{J}$ such that

$$x'Mx = x'A(t)x$$

then $\sup_{t \in \mathfrak{J}} A(t) = M$.

<u>Proof:</u> By (i), M is an upper bound for $\mathfrak{W} = \{A(t), t \in \mathfrak{J}\}$ in $\mathfrak{V}$, the set of n.n.d. matrices of the appropriate dimension. Now suppose that x is an arbitrary vector and that M_1 is any other upper bound for $\mathfrak{W}$ in $\mathfrak{V}$. Then one of the following must occur:

(a) $Mx = 0$, in which case, $x'Mx = 0 \leq x'M_1x$, since $M_1 \in \mathfrak{V}$,

(b) there exists $c_1 \neq 0$ such that $c_1x \in E_M$ which implies, using condition (ii), that there exists $t \in \mathfrak{J}$ such that $1 = c_1^2x'Mx = c_1^2x'A(t)x \leq c_1^2x'M_1x$, where the inequality holds since M_1 is an upper bound for $\mathfrak{W}$.

By (a) and (b) we may conclude that $M \leq M_1$, and, hence, that $\sup_{t \in \mathfrak{J}} A(t) = M$.

4. DIRECTIONALLY MINIMAX MEAN SQUARED ERROR ESTIMATION

In this section we shall obtain a linear estimator of β in the general linear model: $y = X\beta + e$, where y is an $(n \times 1)$ random vector, X is an $(n \times m)$ matrix of known constants, β is an $(m \times 1)$ vector of unknown parameters, and e is a random vector with $E(e) = 0$ and $E(ee') = \Sigma\sigma^2$, $|\Sigma| \neq 0$, $0 < \sigma^2 < \infty$. This model will be referred to in the sequel by $(y, X\beta, \Sigma\sigma^2)$. The procedure to be employed for obtaining the estimator is to evaluate the mean squared error matrix at a value of the parameters, β, which in a sense, is least favorable to estimation, and then to find the member of the class of linear estimators which minimizes the resulting expression. This leads to a minimax estimator.

More precisely, we wish to obtain an estimator of $P'\beta$ of the form Ay, where P' is a given $(p \times m)$ matrix and A is a $(p \times n)$ matrix. For this we shall need the mean squared error matrix:

$$M = M(Ay, P'\beta) = E[(Ay - P'\beta)(Ay - P'\beta)']$$
$$= \sigma^2 A\Sigma A' + (P' - AX)\beta\beta'(P' - AX)'.$$

Following the minimax principle, we wish to evaluate M at the value of β which is least favorable to estimation, that is, which maximizes the bias. It is apparent, however, that M is unbounded in β and, therefore, takes its maximum at infinity. A meaningful modification to the principle is as follows. Write $\beta = k\alpha$, where α is the vector of its direction cosines and k is its length. Then for fixed k, M can be maximized by choice of α. The resulting form then can be minimized by choice of A giving the final form of the estimator. It will turn out that the least favorable α (the least favorable direction) is independent of choice of k. The interpretation that can be placed on this is that in the m-dimensional space of β there is a ray emanating from the origin which corresponds to the worst choice of β with respect to mean squared error. The exact location on the ray is determined by k.

Definition 4.1: A linear estimator, $P'\tilde{\beta} = Ay$, is said to be a Directionally Minimax Mean Squared Error (DMMSE) estimator of $P'\beta$ if A is such that it minimizes

$$S(Ay, P'\beta) = \sigma^2 A\Sigma A' + \sup_{\beta \in \mathcal{B}_k} (P' - AX)\beta\beta'(P' - AX)', \qquad (4.1)$$

where $\mathcal{B}_k = \{\beta | \beta = k\alpha, \alpha'\alpha = 1\}$.

Theorem 4.1: For any $(p \times m)$ matrix, F,

$$\sup_{\beta \in \mathcal{B}_k} F\beta\beta'F' = FF'k^2,$$

where $\mathcal{B}_k = \{\beta | \beta = k\alpha, \alpha'\alpha = 1\}$.

Proof: By the Cauchy-Schwarz inequality, for any x, $x'F\beta\beta'F'x \le x'FF'x\ k^2$ and, hence, $F\beta\beta'F' \le FF'k^2$ for all $\beta \in \mathcal{B}_k$. The result will follow from Theorem 3.1 if, for each $x \in E_{k^2 FF'}$, there exists $\beta_o \in \mathcal{B}_k$ such that $x'F\beta_o\beta_o'F'x = 1$. But if we let $\beta_o = cF'x$ and note that $k^2 = \beta_o'\beta_o = c^2 x'FF'x$, we obtain $\beta_o = (x'FF'x)^{-1/2} k\ F'x$. Then $x'F\beta_o\beta_o'F'x = (x'FF'x)^{-1}(x'FF'x)^2 k^2 = (x'FF'x)k^2 = 1$ by choice of x, which completes the proof.

In view of the last theorem the criterion given by (4.1) can be reduced to:

$$S(Ay, P'\beta) = \sigma^2 A\Sigma A' + k^2 (P' - AX)(P' - AX)'. \qquad (4.2)$$

The following theorem gives the value of A which minimizes (4.2) and, hence, the form of the linear DMMSE estimator.

Theorem 4.2: For the linear model $(y, X\beta, \Sigma\sigma^2)$, the linear DMMSE estimator of $P'\beta$ is

$$\widetilde{P'\beta} = P'X'(\Sigma c^{-2} + XX')^{-1}y = P'\tilde{\beta},$$

and $\min_A S(Ay, P'\beta)/k^2 = P'(I + X'\Sigma^{-1}X c^2)P$, where $c^2 = k^2/\sigma^2$.

Proof: Let $E = (\Sigma c^{-2} + XX')$ and $c^2 = k^2/\sigma^2$, then

$$S(Ay, P'\beta)/k^2 = c^{-2}A\Sigma A' + (P' - AX)(P' - AX)'$$

$$= A(c^{-2}\Sigma + XX')A' + P'P - P'X'A' - AXP$$

$$= AEA' - P'X'E^{-1}EA' - AEE^{-1}XP + P'P + P'X'EE^{-1}XP - P'X'E^{-1}XP$$

$$= (A - P'X'E^{-1})E(A - P'X'E^{-1})' + P'(I - X'E^{-1}X)P$$

$$= (A - P'X'E^{-1})E(A - P'X'E^{-1})' + P'(I + X'\Sigma^{-1}X c^2)^{-1}P.$$

The last equality follows from well known formulas. Observe that only the first term in the last expression of the equality depends on A. It is also n.n.d. and takes a minimum at 0 when $A = P'X'E^{-1}$. Thus, $\widetilde{P'\beta} = P'X'(\Sigma c^{-2} + XX')^{-1}y$. By setting $P' = I$ we find $\tilde{\beta} = P'X'(\Sigma c^{-2} + XX')^{-1}y$ and, hence, $\widetilde{P'\beta} = P'\tilde{\beta}$.

Theorem 4.3: For the linear model $(y, X\beta, I\sigma^2)$, the linear DMMSE estimator of β is equivalent to the Ridge regression estimator, $\tilde{\beta}^* = (X'X + Ic^2)^{-1}X'y$, of Hoerl and Kennard (1970).

Proof: The result follows immediately if we apply standard formulas to the form $X'[c^{-2}I + XX']^{-1}$.

REFERENCES

Bunke, O. (1975), "Minimax linear, ridge and shrunken estimators for linear parameters," Mathematische Operationsforschung und Statistik, 6, 697-701.

Gerig, Thomas M. and Zarate-de-Lara, Guillermo P. (1976), "Estimation in linear models using directionally minimax trace mean squared error," Institute of Statistics Mimeograph Series No. 1084, Raleigh, N.C.

Goldstein, M. and Smith, A.F.M. (1974), "Ridge-type estimators for regression analysis," Journal of the Royal Statistical Society, Ser. B, 36, 284-291.

Hoerl, Arthur E. and Kennard, Robert W. (1970), "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, 12, 55-67.

Marquardt, D. W. (1970), "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," Technometrics, 12, 591-612.

Mayer, Lawrence S. and Willke, Thomas A. (1973), "On biased estimation in linear models," Technometrics, 15, 497-508.

Rao, C. Radhakrishna (1976), "Estimation of parameters in a linear model," The Annals of Statistics, 4, 1023-1037.

Toro-Viscarrondo, Carlos and Wallace, T. D. (1968), "A test of the mean square error criterion for restrictions in linear regression," Journal of the American Statistical Association, 63, 558-572.

APPENDIX

The following is the original proof of Theorem 4.1 which was replaced by the simpler one in the text. For convenience and without loss of generality we shall assume that $k^2 = 1$.

<u>Theorem A.1</u>: Let $T = DD'$, where D is a $(t \times r)$ matrix of rank r. Then $T \geq F\beta\beta'F'$ for all $\beta \in \mathcal{B}_1$ if, and only if, $F = DC$ for some $(r \times m)$ matrix C for which the largest characteristic root of $C'C$, $Ch_M(C'C)$, is less than or equal to 1.

<u>Proof</u>: Assume first that

$$DD' \geq F\beta\beta'F' \text{ for all } \beta \in \mathcal{B}_1.$$

This inequality holds, if and only if,

$$x'F\beta\beta'F'x \leq x'DD'x \text{ for all } x \text{ and for all } \beta \in \mathcal{B}_1. \qquad (A.1)$$

Now if $D'x = 0$ we will have from (A.1) that

$$\beta'F'x = 0 \text{ for all } \beta \in \mathcal{B}_1$$

and, hence, $F'x = 0$. Thus,

$$\mathcal{R}^{\perp}(D') \subseteq \mathcal{R}^{\perp}(F'),$$

where $\mathcal{R}(A)$ denotes the vector space generated by the rows of A, therefore,

$$\mathcal{R}(F') \subseteq \mathcal{R}(D')$$

and from this we have that

$$F = DC \text{ for some } C. \qquad (A.2)$$

If, as we have assumed,

$$F\beta\beta'F' \leq DD' \text{ for all } \beta \in \mathcal{B}_1,$$

then by (A.2) we have

$$DC\beta\beta'C'D' \leq DD' \text{ for all } \beta \in \mathcal{B}_1.$$

This last inequality holds if, and only if,

$$x'DC\beta\beta'C'D'x \leq x'DD'x \text{ for all } \beta \in \mathcal{B}_1 \text{ and all } x.$$

Hence, in particular,

$$m'D'DC\beta\beta'C'D'Dm \leq m'D'DD'Dm \text{ for all } \beta \in \mathcal{B}_1 \text{ and all } m.$$

This happens if, and only if,

$$D'DC\beta\beta'C'C'D \leq D'DD'D \text{ for all } \beta \in \mathcal{B}_1,$$

where $|D'D| \neq 0$ and, therefore, the relation holds if, and only if,

$$C\beta\beta'C' \leq I \text{ for all } \beta \in \mathcal{B}_1. \qquad \text{(A.3)}$$

By (A.3) we have that

$$Ch_M(C\beta\beta'C') \leq 1 \text{ for all } \beta \in \mathcal{B}_1$$

which is true if, and only if,

$$\beta'C'C\beta \leq 1 \text{ for all } \beta \in \mathcal{B}_1$$

which in turn is true if, and only if,

$$Ch_M(C'C) \leq 1.$$

Let us now assume that $F = DC$ for some C and $Ch_M(C'C) \leq 1$.

Then,

$$F\beta\beta'F' = DC\beta\beta'C'D'$$

and, by the Cauchy-Schwarz inequality, we will obtain

$$x'DC\beta\beta'C'D'x \leq x'DD'x\beta'C'C\beta \leq x'DD'x \text{ for all } \beta \in \mathcal{B}_1,$$

and all x.

This implies that

$$DC\beta\beta'C'D \leq DD'$$

and, hence, by our assumption we will conclude

$$F\beta\beta'F' \leq DD' = T.$$

Theorem A.2: $F'F$ is an upper bound for $F\beta\beta'F'$, that is,

$$F\beta\beta'F' \leq FF' \text{ for all } \beta \in \mathcal{B}_1.$$

Proof: The proof follows immediately from the Cauchy-Schwarz inequality.

Theorem A.3: If $T = DD' \geq F\beta\beta'F'$ for all $\beta \in \mathcal{B}_1$ then the n.n.d., symmetric matrix FF' is such that

$$FF' \leq DD' = T.$$

Proof: By Theorem A.1 we can write $F = DC$ for some C. Let $CC' = G\Delta G'$, where $GG' = G'G = I$ and Δ is a diagonal matrix, that is, $\Delta = \text{diag}(\delta_i)$.

It follows immediately from Theorems A.1 and A.2 that $\delta_i \leq 1$ for all i. Therefore,

$$\Delta - I \leq 0$$

if, and only if,

$$x'(\Delta - I)x \leq 0 \text{ for all } x.$$

In particular,

$$m'DG(\Delta - I)G'D'm \leq 0 \text{ for all } m.$$

Therefore,

$$m'(DCC'D' - DD')m \leq 0 \text{ for all } m.$$

Thus, we may conclude by Theorem A.1 that

$$FF' \leq DD'.$$

Theorem A.3 states that if $T \in \mathcal{V}$ is an upper bound for $F\beta\beta'F'$, then it satisfies $T \geq FF'$. Then, in view of Theorem A.2, which states that FF' is also an upper bound for $F\beta\beta'F'$,

$$FF' = \sup_{\beta \in \mathcal{B}_1} F\beta\beta'F'.$$