

ESTIMATION IN LINEAR MODELS USING DIRECTIONALLY MINIMAX
MEAN SQUARED ERROR

by

THOMAS M. GERIG and GUILLERMO P. ZARATE-DE-IARA

Institute of Statistics
Mimeograph Series No. 1084
Raleigh, N.C. - August 1976

1. INTRODUCTION

The most commonly used estimators of the regression coefficients in general linear models are the best linear unbiased estimators (or ordinary least squares estimators). By dropping the unbiasedness criterion, linear estimators can be obtained that have smaller variances. In this situation, it is common to adopt some type of mean squared error criterion which balances increased bias against reduced variance. Unfortunately, estimators derived from this type of criterion invariably lead to expressions that involve the parameters themselves and, as such, are useless in practice.

Several authors have proposed classes of biased regression estimators which can be used as alternatives to the ordinary least squares (OLS) estimator. These estimators are studied and compared to OLS estimators with respect to mean squared error. Typically, they are not uniformly better in this respect but in some situations can be dramatically superior. Some of the common methods of biased regression estimation include Hoerl and Kennard's [3] Ridge regression, Marquardt's [4] Generalized Inverse regression, Mayer and Willke's [5] Shrunken OLS regression, and Toro and Wallace's [7] False Restrictions regression. Many others have proposed modifications and extensions to these.

The purpose of this paper is to develop a meaningful criterion which can be used to derive possibly biased estimators which will be superior to OLS with respect to mean squared error. In Section 3, a criterion is proposed which is based on a minimax argument. In the formula for the trace of the mean squared error matrix, the value of

the regression parameters which are, in a sense, the least favorable to estimation (that is, which result in the greatest bias) are substituted for the unknown parameters. By this, it is hoped that the derived estimators will protect the user from the worst case that nature can produce.

In Section 4, this criterion is applied to the class of estimators obtained by computing OLS estimators subject to possibly false restrictions. The criterion is minimized by choice of a prespecified number of independent restrictions. A statistical test is given which can be used to decide whether the resulting estimator is superior in mean squared error to OLS. The resulting estimators turn out to be equivalent to Marquardt's [4] Generalized Inverse estimators.

Section 5 is devoted to the study of the class of Shrinkage estimators defined by Goldstein and Smith [2]. Within this class the optimum estimator turns out to a member of the subclass of Shrunken OLS estimators defined by Mayer and Willke [5].

2. NOTATION

Consider the linear model

$$y = X\beta + e, \quad (2.1)$$

where y is an $(n \times 1)$ vector of observations, X is an $(n \times m)$ matrix of rank $m (\leq n)$ of known constants, β is an $(m \times 1)$ vector of unknown regression parameters, and e is an $(n \times 1)$ vector of unobservable random variables with $E(e) = 0$ and $E(ee') = \Sigma \sigma^2$,

$$|\Sigma| \neq 0.$$

Define

$$S = X'X \quad (2.2)$$

and

$$\hat{\beta} = S^{-1}X'y \quad (2.3)$$

and write the singular value decomposition of X as

$$X = U\Lambda^{1/2}V' \quad (2.4)$$

where U is $(n \times m)$, $\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_m^{1/2})$, $\lambda_1^{1/2} \geq \dots \geq \lambda_m^{1/2} > 0$, V is $(m \times m)$, $U'U = I_m$, $V'V = VV' = I_m$, and I_m is the $(m \times m)$ identity matrix. From this we write

$$S = V\Lambda V' \quad (2.5)$$

Define the partitioned matrices

$$U = (U_1 : U_2) \quad (2.6)$$

and

$$V = (V_1 : V_2) \quad (2.7)$$

where V_1 is $(m \times m - u)$, V_2 is $(m \times u)$, U_1 is $(n \times m - u)$, and U_2 is $(n \times u)$ for some u . Similarly, define

$$\Lambda^{1/2} = \begin{bmatrix} \Lambda_1^{1/2} & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & \Lambda_2^{1/2} \end{bmatrix} \quad (2.8)$$

where $\Lambda_1^{1/2}$ is $[(m - u) \times (m - u)]$ and $\Lambda_2^{1/2}$ is $(u \times u)$ for some u .

We shall be concerned with the simultaneous estimation of the regression parameters, β , using linear functions of the observations $Ay + b$, where A is some $(m \times m)$ matrix and b an $(m \times 1)$ vector. For this we shall need the mean squared error matrix:

$$\begin{aligned} M &= M(Ay + b, \beta) = E[(Ay + b - \beta)(Ay + b - \beta)'] \\ &= AA'\sigma^2 + (b + F\beta)(b + F\beta)', \end{aligned}$$

where $F = I - AX$. Also we shall need the trace of M ,

$$T(Ay + b, \beta) = \text{tr}[M(Ay + b, \beta)].$$

Finally, define

- (i) $\|x\|^2 = x'x$, for any vector x ,
- (ii) $\text{Ch}_M(A)$ is the largest characteristic root of the matrix A ,
- (iii) $C(A)$ is the vector space generated by the columns of A .

3. DIRECTIONALLY MINIMAX TRACE MEAN SQUARED

ERROR ESTIMATORS

The goal of this work is to obtain an estimator of β which does not depend on the parameters themselves and which, in some sense, minimizes the trace mean squared error, $T(Ay + b, \beta)$. Adopting a minimax philosophy leads to attempting to replace β by the value which maximizes $T(Ay + b, \beta)$. This, of course, is not helpful as T is unboundedly increasing with $\|\beta\|$. The modification adopted in this

work is to express the parameters in the form $\beta = k\alpha$, where α are its direction cosines and k its length. Then, for fixed values of k , the expression, $T(Ay + b, k\alpha)$, can be maximized by choice of α . By Theorem 3.1 below, the ultimate choice of α is independent of the value of k . This fact indicates that, emanating from the origin, there is one direction corresponding to the worst choice of β with respect to mean squared error (or equivalently, squared bias). It is from along this ray that the value of β is chosen to maximize T . The exact location on the ray is set by choice of k .

Definition 3.1: A linear estimator $\hat{\beta} = Ay + b$ is said to be directionally minimax trace mean squared error (DMTMSE) estimator of β if A and b are such that they minimize

$$\begin{aligned} S_T(A, b, k) &= \text{Sup}_{\beta \in \beta_k} T(Ay + b, \beta) \\ &= \sigma^2 \text{tr}(A\Sigma A') + \text{Sup}_{\beta \in \beta_k} (b + F\beta)'(b + F\beta) \end{aligned}$$

where $\beta_k = \{\beta \mid \beta = k\alpha, \|\alpha\| = 1\}$.

Theorem 3.1: In the linear model $(y, X\beta, \Sigma\sigma^2)$, $S_T(A, b, k) \geq S_T(A, 0, k)$ for any A , and in particular, the DMTMSE estimator is of the form Ay .

Proof: Either $\|F\beta + b\|^2 \leq \|F\beta\|^2$ or $\|F\beta + b\|^2 \geq \|F\beta\|^2$. Suppose the former holds, then it follows that $\|b\|^2 \leq -2b'F\beta$ and hence, $\|-F\beta + b\|^2 = \|F\beta\|^2 + \|b\|^2 - 2b'F\beta \geq \|F\beta\|^2 + 2\|b\|^2 \geq \|F\beta\|^2$. Thus,

$$\max(\|F\beta + b\|^2, \|-F\beta + b\|^2) \geq \|F\beta\|^2.$$

Now

$$\begin{aligned} \sup_{\beta \in \mathcal{B}_k} \|\mathbb{F}\beta + b\|^2 &= \sup_{\beta \in \mathcal{B}_k} \{\max[\|\mathbb{F}\beta + b\|^2, \|\mathbb{F}\beta - b\|^2]\} \\ &\geq \sup_{\beta \in \mathcal{B}_k} \|\mathbb{F}\beta\|^2, \end{aligned}$$

with equality when $b = 0$. Therefore, $S_T(A, b, k) \geq S_T(A, 0, k)$, for all A and k .

In view of Theorem 3.1, the criterion can be reduced to:

$$\begin{aligned} S_T(Ay, k) &= \sigma^2 \text{tr}(A\Sigma A') + \sup_{\beta \in \mathcal{B}_k} \beta' F' F \beta \\ &= \sigma^2 \text{tr}(A\Sigma A') + k^2 \text{Ch}_M(F'F), \end{aligned} \quad (3.1)$$

where $F = I - AX$.

4. DMTMSE ESTIMATION IN THE CLASS OF OLS ESTIMATORS

COMPUTED SUBJECT TO FALSE RESTRICTIONS

One class of biased linear estimators of regression coefficients that has been proposed is that obtained by computing least squares estimates under sets of false restrictions. These estimators were studied by Toro and Wallace [7]. Their work is primarily concerned with testing whether or not a particular set of false restrictions will lead to estimators with smaller mean squared error. They leave the choice of restrictions up to the experimenter.

For this case, we shall restrict our attention to the linear model $(y, X\beta, I\sigma^2)$ and assume that X is of full rank ($\text{rank}(X) = m$). To obtain the class of estimators we shall impose u possibly false

restrictions on the parameters:

$$R\beta = h ,$$

where R is a $(u \times m)$ matrix of rank $u (< m)$ and h is a $(u \times 1)$ vector. So that the equations are consistent, we shall assume that $h \in C(R)$.

It is well known (see, for example, Pringle and Rayner [6]) that under this setup the estimator is of the form

$$\begin{aligned} \tilde{\beta} = \tilde{\beta}(R, h) = & [I - S^{-1}R'(RS^{-1}R')^{-1}R]S^{-1}X'y \\ & + S^{-1}R'(RS^{-1}R')^{-1}h , \end{aligned}$$

where S is given by (2.2).

Theorem 4.1: In the class of least squares estimators subject to possibly false restrictions for the setup $(y, X\beta, I\sigma^2)$, restrictions $R\beta = 0$ are preferred over $R\beta = h$ with respect to DTMSE estimation.

Proof: The theorem follows from Theorem 3.1 since, for $R\beta = h$, the estimator will be of the form $Ay + b$, where A is independent of h and b equals 0 whenever $h = 0$.

In view of Theorem 4.1, attention will be limited to the class of least squares estimators obtained subject to $R\beta = 0$.

Theorem 4.2: The class of estimators, $\tilde{\beta}(R)$, obtained by least squares subject to constraints $R\beta = 0$, where $R \in \{R | R \text{ is } (u \times m) \text{ of rank } u\}$ is equivalent to that where $R \in \{R | R \text{ is } (u \times m) \text{ of rank } u \text{ and } RS^{-1}R' = I\}$, where S is given by (2.2).

Proof: Since Λ and V given in (2.4) are positive definite, the rows of $\Lambda^{1/2}V$ form a basis for m -Euclidean space and, hence, R can be written as $R = B\Lambda^{1/2}V'$, for some $(u \times m)$ matrix B . Since $u = \text{rank}(R) \leq \text{rank}(B) \leq u$, B must have full row rank. Now,

$$RS^{-1}R' = B\Lambda^{1/2}V'V\Lambda^{-1}V'V\Lambda^{1/2}B' = BB'$$

and BB' is positive definite. Thus, if we let $RS^{-1}R' = GG'$, $|G| \neq 0$, it is easily seen that $\tilde{\beta}(R) = \tilde{\beta}(G^{-1}R)$. Also, if we let $G^{-1}R = \tilde{R}$, we have

$$\tilde{R}S^{-1}\tilde{R}' = G^{-1}GG'G'^{-1} = I.$$

Thus, for every R there exists a corresponding \tilde{R} such that $\tilde{\beta}(R) = \tilde{\beta}(\tilde{R})$ and $\tilde{R}S^{-1}\tilde{R}' = I$.

Using Theorem 4.2 we can respecify the class of estimators of interest as

$$\tilde{\beta} = \tilde{\beta}(R) = (I - S^{-1}R'R)S^{-1}X'y$$

for all R such that $RS^{-1}R' = I$. Within this class we wish to find the optimum with respect to the DMIMSE criterion set out in Section 3. It should be noted that the value of u is assumed to be given and fixed. Also, because of the assumption that R is of full row rank, the ordinary least squares (OLS) estimator is not in the class. However, there exist members of the class that are arbitrarily "close" to the OLS estimator.

Subsections 4.1 and 4.2 below deal with deriving the optimum estimator for $u = 1$ and for general u , respectively, and subsection 4.3 examines some properties of the resulting estimators.

4.1 The Single Constraint Case

The first case we shall consider is when $u = 1$. In this case the set of restrictions being imposed are of the form $r'\beta = 0$, where r is an $(m \times 1)$ vector satisfying $r'S^{-1}r = 1$.

By putting $A = (I - S^{-1}rr')S^{-1}X'$ in (3.1), and since $\text{Ch}_M(F'F) = \text{Ch}_M(rr'S^{-2}rr') = r'rr'S^{-2}r$, we get

$$\begin{aligned} S_T(r,k) &= \sigma^2 \text{tr}[(I - S^{-1}rr')S^{-1}X'XS^{-1}(I - rr'S^{-1})] \\ &\quad + k^2 r'rr'S^{-2}r \\ &= \sigma^2 \text{tr}[S^{-1} - S^{-1}rr'S^{-1} - S^{-1}rr'S^{-1} + S^{-1}rr'S^{-1}rr'S^{-1}] \\ &\quad + k^2 r'rr'S^{-2}r \\ &= \sigma^2 \text{tr}[S^{-1}] + (k^2 r'r - \sigma^2) r'S^{-2}r \\ &= \sigma^2 \sum_{i=1}^m \lambda_i^{-1} + (k^2 r'r - \sigma^2) r'S^{-2}r, \end{aligned}$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ and Λ is defined in (2.4) and (2.5).

The following theorem gives the vector r which minimizes

$S_T(r,k)$.

Theorem 4.3:

$$S_T(r,k) \geq \sigma^2 \sum_{i=1}^{m-1} \lambda_i^{-1} + k^2$$

with equality when $r = \lambda_m^{-1/2} v_m$, where $V = (v_1, v_2, \dots, v_m)$ and V is defined in (2.4). That is, the DTMSE estimator in the class of least

squares estimators computed subject to one possible incorrect restriction is $\tilde{\beta} = (I - V_m V_m') S^{-1} X' y$.

Proof: Let $w = \Lambda^{-1/2} V' r$ so that $r = V \Lambda^{1/2} w$. Also, $r' S^{-1} r = 1$ implies $w' w = 1$. Now

$$\begin{aligned} S_T(r, k) &= S_T(V \Lambda^{1/2} w, k) \\ &= \sigma^2 \sum_{i=1}^m \lambda_i^{-1} + (k^2 \sum_{i=1}^m w_i^2 \lambda_i^2 - \sigma^2) \sum_{i=1}^m w_i^2 \lambda_i^{-1} \\ &= \sigma^2 \sum_{i=1}^m \lambda_i^{-1} + k^2 \left(\sum_{i=1}^m w_i^2 \lambda_i^2 \right) \left(\sum_{i=1}^m w_i^2 \lambda_i^{-1} \right) \\ &\quad - \sigma^2 \sum_{i=1}^m w_i^2 \lambda_i^{-1}. \end{aligned} \quad (4.1)$$

Since $\sum_{i=1}^m w_i^2 = 1$ and since the harmonic mean is always less than or equal to the arithmetic mean:

$$\left(\sum_{i=1}^m w_i^2 \lambda_i^2 \right) \left(\sum_{i=1}^m w_i^2 \lambda_i^{-1} \right) \geq 1. \quad (4.2)$$

Furthermore, since the value of a weighted arithmetic mean is always less than or equal to the largest value being averaged:

$$\sum_{i=1}^m w_i^2 \lambda_i^{-1} \leq \lambda_m^{-1}. \quad (4.3)$$

Using (4.2) and (4.3) in (4.1) we get

$$\begin{aligned} S_T(r, k) &\geq \sigma^2 \sum_{i=1}^m \lambda_i^{-1} + k^2 - \sigma^2 \sum_{i=1}^m w_i^2 \lambda_i^{-1} \\ &\geq \sigma^2 \sum_{i=1}^m \lambda_i^{-1} + k^2 - \sigma^2 \lambda_m^{-1} \\ &= \sigma^2 \sum_{i=1}^{m-1} \lambda_i^{-1} + k^2. \end{aligned}$$

Noting that equality holds in (4.2) whenever $\lambda_i = \lambda$ for all values of i for which $w_i > 0$ and in (4.3) whenever $w_1 = w_2 = \dots = w_{m-1} = 0$, $w_m = 1$ we see that both equalities hold when $w' = (0, 0, \dots, 0, 1)$, in which case $r = V\Lambda^{1/2}w = \lambda_m^{1/2}V_m$. Thus,

$$S_T(\lambda_m^{1/2}V_m, k) = \sigma^2 \sum_{i=1}^{m-1} \lambda_i^{-1} + k^2$$

and the resulting estimator becomes

$$\tilde{\beta} = (I - V_m V_m') S^{-1} X' y .$$

The estimator thus derived can be seen to be of the form $\tilde{\beta} = P\hat{\beta}$, where P is an orthogonal projection matrix and $\hat{\beta}$ is the OLS estimator. The estimator can be obtained by projecting $\hat{\beta}$ into the space orthogonal to the characteristic vector corresponding to the smallest characteristic root of $X'X$.

Another interesting feature of this result is the fact that the estimator does not depend on the value of k and, hence, need not be specified before hand. Of course, the resulting mean squared errors will involve k .

4.2 The General Case

For the general case we shall assume that u is a fixed number ($1 \leq u \leq m$). The solution has the appearance of that for $u = 1$ but the proof of the theorem is more involved.

Theorem 4.4: For the class of least squares estimates computed subject to a set of u independently possibly incorrect restrictions,

$$S_T(R, k) \geq \sigma^2 \sum_{i=1}^{m-u} \lambda_i^{-1} + k^2$$

with equality holding whenever $R = \Lambda_2^{1/2} V_2'$, where Λ_2 and V_2 are given (2.7) and (2.8). That is, for this class of estimators the DMIMSE estimator is

$$\tilde{\beta} = (I - V_2 V_2') S^{-1} X' y = V_1 V_1' S^{-1} X y.$$

Proof: We can write $R = (BB')^{-1/2} B \Lambda^{1/2} V'$ where B is any $(u \times m)$ matrix of rank u . From this we have that

$$\begin{aligned} \text{tr}[S^{-1} - S^{-1} R' R S^{-1}] &= \text{tr}[V \Lambda^{-1} V' - V \Lambda^{-1} V' V \Lambda^{1/2} B' (BB')^{-1/2} (BB')^{-1/2} B \Lambda^{1/2} V' V \Lambda^{-1} V'] \\ &= \text{tr}[\Lambda^{-1}] - \text{tr}[\Lambda^{-1/2} P_B \Lambda^{-1/2}] \\ &= \sum_{i=1}^m \lambda_i^{-1} - \sum_{i=1}^m p_{ii} \lambda_i^{-1}, \end{aligned}$$

where $P_B = B'(BB')^{-1}B = ((p_{ij}))$. Since P_B is symmetric idempotent we have that $\sum_{j=1}^m p_{ij}^2 = p_{ii}$ which implies that $p_{ii} \geq 0$. Further, if $p_{ii} > 0$ we have that $0 = \sum_{j \neq i}^m p_{jj}^2 + p_{ii}^2 - p_{ij}^2 \geq p_{ii}^2 - p_{ii} - p_{ii} \Rightarrow p_{ii} \leq 1$. It is well known that $\sum_{i=1}^m p_{ii} = u$. Therefore, $\sum_{i=1}^m p_{ii} \lambda_i^{-1}$ is maximized subject to $0 \leq p_{ii} \leq 1$ and $\sum_{i=1}^m p_{ii} = u$ by choosing $p_{11} = \dots = p_{(m-u)(m-u)} = 0$ and $p_{(m-u+1)(m-u+1)} = \dots = p_m = 1$ giving

$$\sum_{i=1}^m p_{ii} \lambda_i^{-1} = \sum_{i=m-u+1}^m \lambda_i^{-1}.$$

This corresponds to taking $B = [0; I_u]$ and

$$P_B = \begin{bmatrix} 0 & \vdots & 0 \\ \dots & \vdots & \dots \\ 0 & \vdots & I_u \end{bmatrix}.$$

Using this we have

$$\text{tr}[S^{-1} - S^{-1}R'RS^{-1}] \geq \sum_{i=1}^{m-u} \lambda_i^{-1}. \quad (4.4)$$

Next, we have that

$$\begin{aligned} \text{Ch}_M(R'RS^{-1}R'R) &= \text{Ch}_M(V\Lambda^{1/2}P_B\Lambda^{-1/2}\Lambda^{-1/2}P_B\Lambda^{1/2}V') \\ &= \text{Ch}_M(\Lambda^{1/2}P_B\Lambda^{-1/2}\Lambda^{-1/2}P_B\Lambda^{1/2}) = \text{Ch}_M(Q'Q), \end{aligned}$$

where $Q = \Lambda^{1/2}P_B\Lambda^{-1/2}$. Note that $Q^2 = Q$ but is not symmetric. Now suppose that $l \in C(Q)$ then $l = Qm$ for some m and, hence, $Ql = l$. Furthermore, if $l^* = l(l'l)^{1/2}$ then $Ql^* = l^*$ and $\|Ql^*\|^2 = \|l^*\|^2 = 1$.

Therefore

$$\text{Ch}_M(Q'Q) = \sup_{x(x'x=1)} \|Qx\|^2 \geq 1$$

and

$$\text{Ch}_M(R'RS^{-1}R'R) \geq 1. \quad (4.5)$$

Using (4.4) and (4.5) we may conclude that

$$\begin{aligned} \sigma^2 \text{tr}(S^{-1} - S^{-1}R'RS^{-1}) + k^2 \text{Ch}_M(R'RS^{-2}R'R) \\ \geq \sigma^2 \sum_{i=1}^{m-u} \lambda_i^{-1} + k^2. \end{aligned} \quad (4.6)$$

It remains to demonstrate that the lower bound is attainable. Putting $B = (0; I_u)$ we get $R = (BB')^{-1/2}B\Lambda^{1/2}V' = \Lambda_2^{1/2}V_2'$ and substituting this into the left hand side of (4.6) gives

$$\begin{aligned} & \sigma^2 \text{tr}(V_1 \Lambda_1^{-1} V_1') + k^2 \text{Ch}_M(V_2 V_2') \\ & = \sigma^2 \sum_{i=1}^{m-u} \lambda_i^{-1} + k^2 \end{aligned}$$

which is the right hand side of (4.6). Finally, the estimator can take various forms:

$$\tilde{\beta} = (I - V_2 V_2') S^{-1} X' y = (I - V_2 V_2') \hat{\beta} = V_1 V_1' \hat{\beta} ,$$

where $\hat{\beta}$ is the OLS estimator.

4.3 Some Properties of the Estimator

Marquardt [4] proposed a class of regression estimators he called Generalized Inverse estimators. He argued that, as an alternative to using precise computing methods for obtaining "exact" solutions in situations where the $X'X$ matrix has some small but positive characteristic roots, it might be preferable to assign a lower rank to $X'X$ and obtain a solution under this assumption. In our notation, his estimator can be written as

$$\hat{\beta}^+ = A_v^+ X' y ,$$

where v is the assigned rank of $X'X$, $A_v^+ = V_1 \Lambda_1^{-1} V_1'$, and V_1 and Λ_1 are defined in (2.7) and (2.8).

Theorem 4.5: The DMTMSE estimator for the class of least squares estimators computed subject to possibly false restrictions is equivalent to Marquardt's Generalized Inverse estimators when the assigned rank of $X'X$ equals $m - u$.

Proof: Since we are assuming that $X'X$ is of full rank,

$$\begin{aligned}\hat{\beta}^+ &= A_V^+ X'y = V_1 \Lambda_1^{-1} V_1' (V_1 \Lambda_1 V_1' + V_2 \Lambda_2 V_2') S^{-1} X'y \\ &= V_1 V_1' S^{-1} X'y = \tilde{\beta}.\end{aligned}$$

Some further results which follow easily from previous results are summarized without proof in the following theorem.

Theorem 4.6: For any $(m \times 1)$ vector l

- (i) $\text{Var}(l'\tilde{\beta}) \leq \text{Var}(l'\hat{\beta})$ with equality if, and only if, $l \in C(V_1)$ in which case $l'\tilde{\beta} = l'\hat{\beta}$.
- (ii) $E(l'\tilde{\beta}) = l'\beta + l'V_2V_2'\beta$ and the second term (the bias) vanishes if, and only if, $l \in C(V_1)$ or $\beta \in C(V_1)$.
- (iii) The estimator $\tilde{\beta}$ is shorter than $\hat{\beta}$, i.e., $\tilde{\beta}'\tilde{\beta} < \hat{\beta}'\hat{\beta}$.
- (iv) For the more general model $(y, X\beta, \Sigma^2)$, where $|\Sigma| \neq 0$, the corresponding estimator is of the form

$$\tilde{\beta} = \tilde{V}_1 \tilde{V}_1' (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}y,$$

where $X'\Sigma^{-1}X = \tilde{V}\tilde{\Lambda}\tilde{V}'$, $\tilde{V}\tilde{V}' = \tilde{V}'\tilde{V} = I$, $\tilde{\Lambda}$ is diagonal,

$\tilde{V} = (\tilde{V}_1 : \tilde{V}_2)$, and \tilde{V}_1 is a $(m \times m - u)$ matrix.

Toro and Wallace [7] give a statistical test for determining whether imposing a given set of possibly false restrictions, $R\beta = 0$, will result in a reduced mean square error. This test is appropriate for use in this setting provided we make the assumption that the errors in the model are normally distributed. In this case, the form of the test is

Reject H_0 if $W \geq W_\alpha$,

where $W = [y'U_2U_2'y/u]/[y'(I - UU')y/(n-m)]$ and U and U_2 are defined in (2.4) and (2.6). Note that the hypothesis H_0 states that $\tilde{\beta}$ is superior to the OLS estimator, $\hat{\beta}$, with respect to mean squared error and accepting H_0 suggests that $\tilde{\beta}$ is preferred over $\hat{\beta}$. It can be seen that W is a noncentral F random variable with u and $(n - m)$ degrees of freedom and noncentrality $\delta = \beta'V_2\Lambda_2V_2'\beta/2\sigma^2$, where V_2 and Λ_2 are defined in (2.7) and (2.8). It can be seen also that $\delta \leq \beta'V_2V_2'\beta\lambda_m/2\sigma^2$, where λ_m is defined in (2.4) and (2.5). Therefore, δ is small whenever the projection of β on $C(V_2)$ is small or λ_m is small. The latter condition is precisely that for which the Generalized Inverse estimators are intended. Some critical points and power computations are given in [7].

5. THE DMTMSE ESTIMATOR FOR THE CLASS OF SHRINKAGE ESTIMATORS

Many regression estimators have been proposed that have the form $\tilde{\beta} = Ay = V\Gamma U'y$, where U and V are defined by (2.4) and $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_m)$. Included in this class are OLS ($\gamma_j = \lambda_j^{-1/2}$, $j = 1, \dots, m$), Generalized Inverse regression ($\gamma_j = \lambda_j^{-1/2}$, $j = 1, \dots, m - u$, $\gamma_j = 0$, $j = m - u + 1, \dots, m$), Ridge regression ($\gamma_j = \lambda_j^{1/2}/(\lambda_j + k^2)$, $j = 1, \dots, m$), as well as others. This class is discussed by Goldstein and Smith [2].

In the following theorem, the DMTMSE estimator for this class is derived.

Theorem 5.1: In the linear model $(y, X\beta, \sigma^2)$, the DMTMSE estimator from the class of estimators of the form $\tilde{\beta} = V\Gamma U'y$, where U and V are defined by (2.4) and $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_m)$, is $\tilde{\beta} = t\hat{\beta}$, where $\hat{\beta}$ is the OLS estimator, $t = c^2 / (\sum_{j=1}^m \lambda_j^{-1} + c^2)$ and $c^2 = k^2 / \sigma^2$.

Proof: First,

$$\begin{aligned} S_T(Ay, k) &= \sigma^2 \text{tr}(AA') + k^2 \text{Ch}_M[(I - AX)(I - X'A')] \\ &= \sigma^2 \sum_{j=1}^m \gamma_j^2 + k^2 \text{Ch}_M(V(I - \Lambda^{1/2} \Gamma)^2 V') \\ &= \sigma^2 \sum_{j=1}^m \gamma_j^2 + k^2 \max[(1 - \lambda_j^{1/2} \gamma_j)^2, j = 1, 2, \dots, m]. \end{aligned}$$

Suppose that the value of the second term is p^2 . Then any set of γ_j 's that minimizes S_T must satisfy $(1 - \lambda_j^{1/2} \gamma_j)^2 \leq p^2$ for $j = 1, \dots, m$ and, hence, $\gamma_j \geq (1-p)\lambda_j^{-1/2}$, $j = 1, \dots, m$. But S_T is a minimum, therefore γ_j must equal $(1-p)\lambda_j^{-1/2}$, its smallest possible value, so as to make $\sum_{j=1}^m \gamma_j^2$ as small as possible. This implies that $S_T / \sigma^2 = (1-p)^2 \sum_{j=1}^m \lambda_j^{-1} + p^2 c^2$. Since this is concave in p , the minimum can be found by setting first derivatives equal to zero resulting in

$$p = \frac{\sum_{j=1}^m \lambda_j^{-1}}{\sum_{j=1}^m \lambda_j^{-1} + c^2}.$$

Thus,

$$\gamma_j = [c^2 / (\sum_{j=1}^m \lambda_j^{-1} + c^2)] \lambda_j^{-1/2}$$

where $c^2 = k^2 / \sigma^2$.

The estimator obtained in Theorem 5.1 is a deterministically shrunken estimator as defined and studied in Mayer and Willke [5]. Its form is simply a scalar times the OLS estimator. The scalar factor is between 0 and 1 and, as such, has the effect of shortening the length of the vector $\hat{\beta}$.

6. THE DMMSE ESTIMATOR FOR THE CLASS OF GENERAL LINEAR FUNCTIONS

The model

$$y = X\beta + e = U\Lambda^{1/2}V'\beta + e$$

can be decomposed as follows. Let \tilde{U} an $(n \times n - m)$ matrix of rank $n - m$ which satisfies $\tilde{U}'U = 0$ and $\tilde{U}'\tilde{U} = I_{n-m}$. That is, let \tilde{U} be the orthogonal complement of U . Then

$$\begin{bmatrix} U' \\ \vdots \\ \tilde{U}' \end{bmatrix} y = \begin{bmatrix} y_1 \\ \vdots \\ y_2 \end{bmatrix} = \begin{bmatrix} \Lambda^{1/2}V' \\ \vdots \\ 0 \end{bmatrix} \beta + \begin{bmatrix} e_1 \\ \vdots \\ e_2 \end{bmatrix}$$

where $E(e_i) = 0$, $\text{Var}(e_i) = I\sigma^2$, and $\text{Cov}(e_1, e_2) = 0$. Thus, it seems sensible to base the estimation of β solely upon y_1 . For purposes of estimating β , then, we shall rewrite the model as:

$$y_1 = \Lambda^{1/2}V'\beta + e_1.$$

It can also be argued that, since V is positive definite, linear estimation of β is equivalent to the estimation of $\beta_1 = \Lambda^{1/2}V'\beta$.

Using this, the model becomes

$$y_1 = \beta_1 + e_1.$$

(6.1)

Theorem 5.1: The DMTMSE linear estimator of β_1 in (5.1) is given by

$$\tilde{\beta}_1 = t y_1, \text{ where } t = c^2/(m + c^2) \text{ and } y_1 \text{ is defined in (6.1).}$$

Proof: (Cohen [1]). Since $X = I$, we have

$$S_T(Ay_1, \beta_1) = \sigma^2 \text{tr}(AA') + k^2 \text{Ch}_M((I-A)(I-A')) .$$

Suppose that A is not symmetric, then define $A^* = I - [(I-A)(I-A')]^{1/2}$

which is symmetric. Note first that $\text{Ch}_M[(I-A)(I-A')] =$

$\text{Ch}_M[(I-A^*)(I-A^{*'})]$. Also $\text{tr}[(I-A)(I-A')] = \text{tr}[(I-A^*)(I-A^{*'})]$. Thus,

$\text{tr}(I) - 2\text{tr}(A) + \text{tr}(AA') = \text{tr}(I) - 2\text{tr}(A^*) + \text{tr}(A^*A^{*'})$. Therefore,

$\text{tr}(A) \geq \text{tr}(A^*)$ implies $\text{tr}(AA') \geq \text{tr}(A^*A^{*'})$. But $\text{tr}(I-A^*) =$

$\text{tr}\{[(I-A)(I-A')]^{1/2}\} \geq \text{tr}(I-A)$, which implies $\text{tr}(A) \geq \text{tr}(A^*)$. Thus,

$S_T(Ay_1, \beta_1) \geq S_T(A^*y_1, \beta_1)$. Assuming that A is symmetric, we write

$A = G\Gamma G'$ and

$$\begin{aligned} S_T(Ay_1, \beta_1) &= \sigma^2 \sum_{j=1}^m \gamma_j^2 + k^2 \text{Ch}_M(G(I - \Gamma^2)G') \\ &= \sigma^2 \sum_{j=1}^n \gamma_j^2 + k^2 \max((1 - \gamma_j^2), j = 1, \dots, m) . \end{aligned}$$

Using the identical argument of Theorem 5.1, the minimizing value of

$\gamma_j = c^2/(m + c^2)$, $j = 1, \dots, m$. The choice of G is arbitrary.

Using the result of Theorem 6.1, an estimator for β can be obtained in terms of y :

$$\tilde{\beta} = v\Lambda^{-1/2}\tilde{\beta}_1 = tv\Lambda^{-1/2}y_1 = tv\Lambda^{-1/2}U'y = t\hat{\beta}$$

where $t = c^2/(m + c^2)$. This is a solution for the class of general linear estimators in the decomposed case given by (6.1). However, because the DMTMSE estimator of $C\beta$ is not necessarily equal to C times

that of β alone, it is not the solution for the original model. Indeed, the form of the estimator in the present case is $V\Lambda^{-1/2}HU'y$ for some H .

REFERENCES

- [1] Cohen, Arthur, "All admissible linear estimators of the mean vector," Annals of Mathematical Statistics, 37 (April 1966), 458-463.
- [2] Goldstein, M. and Smith, A.F.M., "Ridge-type estimators for regression analysis," Journal of the Royal Statistical Society, Ser. B, 36 (June 1974), 284-291.
- [3] Hoerl, Arthur E. and Kennard, Robert W., "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, 12 (February 1970), 55-67.
- [4] Marquardt, D. W., "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," Technometrics, 72 (August 1970), 591-612.
- [5] Mayer, Lawrence S. and Willke, Thomas A., "On biased estimation in linear models," Technometrics, 15 (August 1973), 497-508.
- [6] Pringle, R. M. and Rayner, A. A., Generalized Inverse Matrices With Applications to Statistics, New York: Hafner Publishing Co., 1971.
- [7] Toro-Viscarrondo, Carlos and Wallace, T.D., "A test of the mean square error criterion for restrictions in linear regression," Journal of the American Statistical Association, 63 (June 1968), 558-572.

APPENDIX

Alternative proof of Theorem 6.1:

Since $X = I$, we have

$$S_T(Ay_1, \beta_1) = \sigma^2 \text{tr}(AA') + k^2 \text{Ch}_M((I-A)(I-A')) \quad (1)$$

Write $A = G\Gamma H'$, where G and H are unitary matrices, therefore

$$\text{tr}(AA') = \text{tr}[G\Gamma H' H \Gamma G'] = \text{tr}[\Gamma^2] = \sum_{i=1}^m \gamma_i^2$$

so that the trace term is independent of G and H . Note now that

$$\begin{aligned} \text{Ch}_M((I-A)(I-A')) &= \text{Ch}_M[(I-G\Gamma H')(I-H\Gamma G')] \\ &= \text{Ch}_M[(G'H-\Gamma)'(G'H-\Gamma)'] = \\ &= \text{Ch}_M[(U-\Gamma)(U-\Gamma)] \\ &= \|(U-\Gamma)\|^2 = \|I-U'\Gamma\|^2, \end{aligned} \quad (2)$$

where $U = G'H$, and $\|\cdot\|$ denote the matrix norm induced by the euclidean norm (known as the Spectral Norm)*.

Take γ_i such that $\|I-\Gamma\| = |1-\gamma_i|$, since $\Gamma e_i = \gamma_i e_i$ where $e_i' = [0, \dots, 0, 1, 0, \dots, 0]$, it follows that

$$\begin{aligned} (I - U'\Gamma)e_i &= e_i - U'\Gamma e_i = e_i - \gamma_i U'e_i \\ &= e_i - \gamma_i(\alpha_1 e_i + c) = (1-\gamma_i \alpha_1)e_i - \gamma_i c \end{aligned}$$

* See Lancaster Peter, Theory of Matrices, Academic Press, New York and London, page 210.

where $\alpha_1^2 + \|c\|^2 = 1$ and c is orthogonal to e_i . Hence

$$\begin{aligned} \|I - U'\Gamma\|^2 &\geq (1 - \gamma_i \alpha_1)^2 + \gamma_i^2 \|c\|^2 = 1 \pm \gamma_i^2 (\alpha_1^2 + \|c\|^2) - 2\gamma_i \alpha_1 \\ &\geq 1 + \gamma_i^2 - 2\gamma_i = (1 - \gamma_i)^2 = \|I - \Gamma\|^2 \end{aligned} \quad (3)$$

by applying (3) in (2) we conclude

$$\text{Ch}_M[(U - \Gamma)'(U - \Gamma)] \geq \text{Ch}_M((I - \Gamma)'(I - \Gamma))$$

and the equality is attained letting $G'H = U = I$ which implies that $G = H$, hence, we write $A = G\Gamma G'$ and

$$\begin{aligned} S_T(Ay_1, \beta_1) &= \sigma^2 \sum_{j=1}^m \gamma_j^2 + k^2 \text{Ch}_M(G(I - \Gamma^2)G') \\ &= \sigma^2 \sum_{j=1}^m \gamma_j^2 + k^2 \max[(1 - \gamma_j^2), j = 1, \dots, m] \end{aligned}$$

using the identical argument of Theorem 5.1, the minimizing value of

$$\gamma_j = \frac{c^2}{m + c^2}, \quad j = 1, \dots, m.$$