Some Non-Traditional Minimax

and Sometimes Admissible Estimators

by

M. E. Bock and G. G. Judge

Department of Statistics
Division of Mathematical Sciences
Mimeo Series #79-24

July 18, 1977

Some Non-Traditional Minimax

and Sometimes Admissible Estimators

by

M. E. Bock and G. G. Judge

1. INTRODUCTION

It would seem that if econometric models are constructed and estimated as a source of information for decision making or choice, statistical decision theory, based on the analysis of losses due to incorrect decisions, can and should be used. Consequently, in this paper we make use of decision theoretic procedures as a basis for gauging the sampling performance of alternative estimators (decision rules) for a range of econometric models.

Within this context the widely used least squares estimator for estimating the unknown coefficient vector in the general linear statistical model is minimax and minimum variance unbiased when the random errors have the usual normal distribution. The discovery of the "Stein effect" [12] brought out the fact that the least squares estimator is inadmissible under squared error loss when the size of the coefficient vector is three or more and James and Stein [9] specified an estimator that demonstrated this inadmissibility. Since this development users of the linear model have been presented with a bewildering array of new estimators for the coefficient vector [8], all of which dominate the traditional least squares estimator. These new estimators are all minimax and biased, a few of them are admissible. Furthermore their risk functions are usually highly dependent on the design matrix in the linear model and the specified design matrix may preclude consideration of certain forms of estimators. In this paper we develop a general minimax estimator, discuss its general form and note that it includes many of the ridge-type and Stein-type estimators.

2. MODEL

Assume the following linear statistical model

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

where y is a (T x 1) vector of observations, X is a T x K nonstochastic known matrix of rank $K \geq 3$, β is a K x 1 vector of unknown coefficients, and the unobserved random vector e has a multivariate normal distribution with mean vector 0 and $E[ee'] = \sigma^2 I_T$, with unknown scalar $\sigma^2 > 0$.

We consider estimators $\hat{\beta}$ of the unknown vector β under the risk

(2.2) $$\rho(\beta,\sigma^2;\hat{\beta}) = E[(\hat{\beta}-\beta)'Q(\hat{\beta}-\beta)]/\sigma^2 ,$$

where Q is a K x K specified positive definite matrix.

The least squares estimator of β,

(2.3) $$b = (X'X)^{-1}X'y ,$$

is unbiased and under a squared error loss measure is minimax, and $\rho(\beta,\sigma^2;b) = \text{trace } Q(X'X)^{-1}$. Estimators of σ^2 are often functions of

(2.4) $$s = (y-Xb)'(y-Xb) ,$$

which is independent of b.

Note that because the risk of the maximum likelihood estimator b is constant for all values of the unknown parameters, any estimator $\hat{\beta}$ of β which is distinct from b and is also minimax must have risk less than or equal to that of b everywhere, i.e., $\rho(\beta,\sigma^2;\hat{\beta}) \leq \rho(\beta,\sigma^2;b)$ for all values of β and $\sigma^2 > 0$. If the above inequality holds for some value of the parameters we say that $\hat{\beta}$ dominates b. If one estimator dominates a second estimator we say the second estimator is inadmissible. The admissible estimators for this problem are all generalized Bayes. Not all generalized Bayes estimators are admissible, but all the proper Bayes estimators are admissible [10].

There are of course various possible specifications or forms of uncertainty which may be imposed on the statistical model. Because of this range of alternative statistical models a number of estimators have arisen in the literature and their properties and relationship to one another and to the maximum likelihood estimator are often unclear. For example, one possible specification involves certain elements in the β vector being a priori set equal to fixed values, say zero, and preliminary test estimators [5] have arisen out of this format. Alternatively another possible specification is that of a certain prior distribution for all or part of the elements of the unknown coefficient vector β. Uncertainty as to the exact form of the prior density function has led to a consideration of empirical Bayes [8] and Stein-type estimators.

3. A FAMILY OF ESTIMATORS

We will examine estimators which "correct" the least squares estimator by some matrix H whose stochastic elements are functions of the quadratic form $\underline{b}'Bb/s$ where B is a positive definite matrix.

Let

(3.1) $$\delta(b,s) = (I-H)b$$

where $Q^{\frac{1}{2}}HQ^{-\frac{1}{2}}$ is a positive definite matrix which commutes with the positive definite matrices $Q^{-\frac{1}{2}}BQ^{-\frac{1}{2}}$ and $Q^{\frac{1}{2}}(X'X)^{-1}Q^{\frac{1}{2}}$, which also commute with each other. It is assumed that there is an orthogonal nonstochastic matrix P such that $PQ^{-\frac{1}{2}}BQ^{-\frac{1}{2}}P'$, $PQ^{\frac{1}{2}}HQ^{-\frac{1}{2}}P'$ and $PQ^{\frac{1}{2}}(X'X)^{-1}Q^{\frac{1}{2}}P'$ are diagonal matrices with diagonal elements f_i, h_i and d_i^{-1} respectively. The characteristic roots h_i of $Q^{\frac{1}{2}}HQ^{-\frac{1}{2}}$ are assumed to be functions of $(b'Bb/s)$. If any of the h_i's take on values greater than one, the estimator δ may be improved upon by a "positive part" counterpart [2],

$$\delta^+(b,s) = (I-H^+)b \quad . \tag{3.2}$$

The matrix H^+ commutes with $Q^{-\frac{1}{2}}BQ^{-\frac{1}{2}}$ and $Q^{\frac{1}{2}}(X'X)^{-1}Q^{\frac{1}{2}}$ and is chosen such that the matrix $PQ^{\frac{1}{2}}H^+Q^{-\frac{1}{2}}P'$ is diagonal with elements

$$h_i^+ = \begin{cases} h_i & \text{if } h_i \leq 1 \\ \alpha_i & \text{if } h_i > 1 \end{cases} , \tag{3.3}$$

where α_i is any real-valued measurable function of $b'Bb/s$ with $2-h_i \leq \alpha_i \leq 1$ for $h_i > 1$. So $\rho(\beta,\sigma^2;\delta^+) \leq \rho(\beta,\sigma^2,\delta)$, with strict inequality if δ and δ^+ differ on a set of positive measure.

As an aside it should be noted that a number of ridge type estimators [6] have been considered which have the above form with

$$h_i = 1/(1+\rho_i) \quad , \tag{3.4}$$

where ρ_i is a function of a quadratic form in b. For example, if we set $\sigma^2 = 1$ and set $Q = I$ with $a_i \geq 0$, and $0 \leq r^{-1}(\cdot) \leq 1$ where d_i^{-1} is the ith characteristic root of $(X'X)^{-1}$, results of Casella [6] show that if certain regularity conditions on r are satisfied and if

$$\rho_i = b'X'Xb \; a_i \; d_i \, r\,(b'X'Xb) \tag{3.5}$$

and

(3.6) $$\max \{a_i^{-1} d_i^{-1}\} \leq 2\ [(\Sigma\ a_i^{-1}\ d_i^{-2}/\max\ \{a_i^{-1}\ d_i^{-2}\}) - 2],$$

then the ridge estimator is minimax. If $a_i = 1$, then tr $(X'X)^{-2}$ must be greater than or equal to 2 max. ch. rt. $(X'X)^{-2}$. In general the a_i may be chosen such that (3.6) is satisfied.

Furthermore a number of priors may be specified for which the resulting Bayes estimator has the form of (3.1) and is minimax.

General Minimax Estimator

Assume that the matrix H has the form

(3.7) $$H = h(b'Bb/s)C\ ,$$

where C is a known matrix and h is a real-valued measurable function of $b'Bb/s$. Our task is to give conditions on h which insure that the estimator

(3.8) $$\delta(b,s) = [I_k - h(b'Bb/s)C]b$$

is minimax.

At the outset note that for certain design matrices X and certain specifications of B and C, no function h exists for which δ is minimax. In particular, if $B = (X'X)$ and $C = I$ and $Q = I$, then no estimator of the form δ is minimax if $tr(X'X)^{-1}$ is less than twice the maximum characteristic root (max ch rt) of $(X'X)^{-1}$ [4]. For the regression problem, certain Stein-type estimators do not dominate the least squares estimator as noted in the comments to the paper by Dempster, Schatzoff and Wermuth [7]. The conditions to be given for the minimaxity of δ are generalizations of those developed by Efron and Morris [8] for the case when $(X'X) = I_K$.

Theorem: Assume $Q^{\frac{1}{2}}CQ^{-\frac{1}{2}}$ and $Q^{-\frac{1}{2}}BQ^{-\frac{1}{2}}$ are positive definite matrices which commute with each other and with $Q^{\frac{1}{2}}(X'X)^{-1}Q^{\frac{1}{2}}$. Then

$$(3.8) \qquad \delta(b,s) = \left[I_K - h\left(\frac{b'Bb}{s}\right) C\right] b$$

is minimax if the following conditions hold:

$$(i) \qquad 0 < c_o = \frac{2[tr(C(X'X)^{-1}Q) - 2 \max ch\ rt \{C(X'X)^{-1}Q\}]}{(T-K+2)(\max ch\ rt\{C'QCB^{-1}\})},$$

(which implies $K > 2$);

$$(ii) \qquad 0 < h(u) < \frac{c_o}{u},$$

for all $u > 0$ and h is differentiable for all $u > 0$; and

$$(iii) \qquad \psi_1(u) = u^{d_o}\left[\frac{c_o}{u} - h(u)\right]^{(1+f_o)} h(u)$$

is non-decreasing function of u if $h(u) < c_o/u$, where

$$d_o = \left[\frac{(T-K-2)}{4} c_o \frac{\max ch\ rt \{C'QCB^{-1}\}}{\max ch\ rt \{C(X'X)^{-1}Q\}}\right] \text{ and } f_o = \left[(4/(T-K-2)d_o\right].$$

Furthermore if $h(u_o) = c_o/u_o$, assume $h(u) = c_o/u$ for $u > u_o$.

The proof of the general minimax theorem is as follows: Because $Q^{\frac{1}{2}}CQ^{-\frac{1}{2}}$, $Q^{-\frac{1}{2}}BQ^{-\frac{1}{2}}$ and $Q^{\frac{1}{2}}(X'X)^{-1}Q^{\frac{1}{2}}$ commute, there is an orthogonal matrix P such that the matrices are simultaneously diagonalizable as previously noted with $h_i = h(b'Bb/s)c_i$.

Setting $w = \frac{(PQ^{+\frac{1}{2}}b)}{\sigma}$, we have that

(3.9)
$$\rho(\beta,\sigma^2;\delta) = E\left[\frac{(b-\beta)'Q(b-\beta)}{\sigma^2}\right] + \left[E\, h^2\left(s^{-1}b'Bb\right)\frac{b'C'QCb}{\sigma^2}\right] - 2E\left[\frac{(b-\beta)'}{\sigma^2}QCB\, h\left(s^{-1}b'Bb\right)\right]$$
$$= tr(X'X)^{-1}Q + E\left[h^2\left(\frac{\Sigma f_i w_i^2}{\chi^2_{(T-K)}}\right)\Sigma c_i^2 w_i^2\right] - 2\sum_{i=1}^{K} c_i d_i^{-1}$$
$$E\left[(w_i - E[w_i])d_i\, h\left(\frac{\Sigma f_j w_j^2}{\chi^2_{(T-K)}}\right)w_i\right]$$

where $\chi^2_{(T-K)}$ has the chi-square distribution with (T-K) degrees of freedom and is independent of w.

Differentiation and integration by parts of $\left[h\left(\frac{\Sigma f_j w_j^2}{\chi^2_{(T-K)}}\right) w_i \exp\left(\frac{\Sigma d_j}{2}(w_j - E[w_j])^2\right)\right]$ implies that

(3.10)
$$E[(w_i - E[w_i])d_i\, h(\Sigma f_j w_j^2/\chi^2_{(T-K)})w_i] = E\left[\frac{\partial}{\partial w_i}\{h(\Sigma f_j w_j^2/\chi^2_{(T-K)})w_i\}\right].$$

Substituting

(3.11)
$$E\left[\frac{\partial}{\partial w_i}\left[h\left(\frac{\Sigma f_j w_j^2}{\chi^2_{(T-K)}}\right)w_i\right]\right] = E\left[h\left(\frac{\Sigma f_j w_j^2}{\chi^2_{(T-K)}}\right)\right] + 2E\left[\frac{f_i w_i^2}{\chi^2_{(T-K)}}\quad h'\left(\frac{\Sigma f_j w_j^2}{\chi^2_{(T-K)}}\right)\right]$$

for $\left[E\left[(w_i - Ew_i)d_i h\left(\frac{\Sigma f_j w_j^2}{\chi^2_{(T-K)}}\right)\right]\right]$ in $\rho(\beta,\sigma^2;\delta)$ yields for $u = \Sigma f_i w_i^2/\chi^2_{(T-K)}$,

(3.12)
$$\rho(\beta,\sigma^2;\delta) = tr(X'X)^{-1}Q + E\left[h^2(u)\left(\frac{\Sigma c_i^2 w_i^2}{\chi^2_{(T-K)}}\right)\chi^2_{(T-K)}\right]$$
$$- 2\left(\sum_{i=1}^{K} c_i d_i^{-1}\right)E\left[h\ (u)\right] - 4E\left[\frac{\Sigma c_i d_i^{-1} f_i w_i^2}{\chi^2_{(T-K)}}h'\ (u)\right].$$

Note in line with [8] that $\left[E[g(\chi^2_{(n)})\chi^2_{(n)}] = nE[g(\chi^2_{(n)})] + 2E[g'(\chi^2_{(n)})\chi^2_{(n)}]\right]$.

Setting $\left[g(\chi^2_{(n)}) = h^2\left(\frac{\Sigma f_i w_i^2}{\chi^2_{(n)}}\right)\frac{\Sigma c_i^2 w_i^2}{\chi^2_{(n)}}\right]$, and applying this result implies

$$(3.13)\quad E\left[h^2(u)\left(\frac{\Sigma c_i^2 w_i^2}{\chi^2_{(T-K)}}\right)\chi^2_{(T-K)}\right] = (T-K-2)\,E\left[h^2(u)\left(\frac{\Sigma c_i^2 w_i^2}{\chi^2_{(T-K)}}\right)\right] - 4E\left[h'(u)h(u)\left(\frac{\Sigma c_i^2 w_i^2}{\chi^2_{(T-K)}}\right)\right].$$

Substituting this in the last expression for $\rho(\beta,\sigma^2;\delta)$ we have

$$(3.14)\quad \rho(\beta,\sigma^2;\delta) = \mathrm{tr}\,(X'X)^{-1}Q + E\Bigg[h(u)\,u\left((T-K-2)\,h(u)\left(\frac{\Sigma c_i^2 w_i^2}{\Sigma f_i w_i^2}\right) - \frac{2(\Sigma d^{-1} c_i)}{u}\right)$$

$$- 4h'(u)\,u\left(\left(\frac{\Sigma c_i d_i^{-1} f_i^2 w_i^2}{\Sigma f_i w_i^2}\right) + \left(\frac{\Sigma c_i^2 w_i^2}{\chi^2_{(T-K)}}\right)h(u)\right)\Bigg]$$

$$\leq \mathrm{tr}(X'X)^{-1}Q + E\Bigg[\left(\frac{\Sigma c_i d_i^{-1} f_i w_i^2}{f_i w_i^2} + \frac{\Sigma c_i^2 w_i^2}{\chi^2_{(T-K)}}\,h(u)\right)\left(1 + (f_o/c_o)\,h(u)\,u\right)^{-1}$$

$$\left(h(u)\,u\,(\max_i\{c_i d_i^{-1}\})^{-1}\left((T-K-2)\,h(u)\max_i\{c_i^2 f_i^{-1}\} - \frac{2(\Sigma c_i d_i^{-1})}{u}\right)\right.$$

$$\left. - 4h'(u)\,u\,(1 + (f_o/c_o)\,h(u)\,u)\right)\Bigg].$$

This last inequality is true because

$$(3.15)\qquad h(u)\,(T-K-2)\left(\frac{\Sigma c_i^2 w_i^2}{\Sigma f_i w_i^2}\right) < h(u)\,(T-K-2)\max_i\{c_i^2 f_i^{-1}\}$$

and

$$(3.16)\qquad [(T-K-2)\,h(u)\max_i\{c_i^2 f_i^{-1}\} - 2(c_i d_i^{-1})/u]$$

is non-positive, and

$$(3.17)\qquad \left(\left(\frac{\Sigma c_i d_i^{-1} f_i w_i^2}{\Sigma f_i w_i^2}\right) + \left(\frac{\Sigma c_i^2 w_i^2}{\chi^2_{(T-K)}}\right)h(u)\right)\left(1 + (f_o/c_o)\,h(u)\,u\right)^{-1} \leq \max_i\{c_i d_i^{-1}\}.$$

Using the definition of ψ_1 we may rewrite the last upper bound so that

$$(3.18)\qquad \rho(\beta,\sigma^2;\delta) < \mathrm{tr}(X'X)^{-1}Q - E\left[4\left(\left(\frac{\Sigma c_i d_i^{-1} f_i w_i^2}{\Sigma f_i w_i^2}\right)+\left(\frac{\Sigma c_i^2 w_i^2}{\chi^2_{(T-K)}}\right)h(u)\right)(1+(f_o/c_o)\; h(u)\; u)^{-1}\psi_1'(u)\; c_o^{-1} u^{2-d_o} (c_o/u - h(u))^{2+f_o}\right]$$

$$< \mathrm{tr}(X'X)^{-1}Q, \text{ if } \psi_1'(\cdot) > 0 ,$$

where

$$(3.19)\quad \psi_1'(u) = -c_o u^{(d_o-2)} (c_o/u - h(u))^{-(2+f_o)}\left[(d_o/c_o)\, u\, h^2(u) - u\, h'(u)\left(1 + (f_o/c_o)\, u\, h(u)\right) - h(u)(d_o+f_o+1)\right]$$

if $h(u) \neq c_o/u$. If $h(u_o) = c_o/u_o$, then $h(u) = c_o/u$ for $u > u_o$, and we define $\psi_1'(u) = 0$ for $u > u_o$. Since $\psi_1' > 0$ we have that $\rho(\beta,\sigma^2;\delta) < \mathrm{tr}(X'X)^{-1}Q$ which implies that δ is minimax.

These results imply there exists a large class of estimators that are superior to many of the maximum likelihood estimators employed in econometric work if performance is evaluated under the squared error loss measure.

Admissible Estimators

It has been noted by Berger and Srinivasan [3] that when $Q = I$ and σ^2 is known, a necessary condition for admissibility of estimators of β of the form

$$(3.8)\qquad \delta(b) = (I - h(b'Bb)C)b$$

is that $C = \sigma^2(X'X)^{-1}B$. This would suggest that there may be some form of estimator (as in the case of the positive part estimator) which automatically improves on estimators of the above form if $C \neq \sigma^2(X'X)^{-1}B$. The condition that $C = \sigma^2(X'X)^{-1}B$ arises from a much more general condition for the admissibility of estimators of β that is given by Berger and Srinivasan [3], which says that an admissible estimator δ^* of β must satisfy the condition that $\sigma^{-2}X'X\delta^*(b)$ has a symmetric Jacobian matrix.

Referring back to the minimax condition implied by (1) of Theorem,

$$\text{tr}(C(X'X)Q^{-1}) > 2 \text{ max. ch. rt. } (C(X'X)Q^{-1}) \;, \tag{3.20}$$

it is clear that the choice of $C = Q^{-1}\, X'X$ would insure that (3.20) is true for any $X'X$ matrix and specification of Q. Thus

$$\delta(b) = (I_K - \sigma^2 \; h \; (b'(X'X)^2 b) \; X'X)b \tag{3.21}$$

is one estimator of the above form when $Q = I$ which satisfies the necessary condition for admissibility given by Berger and Srinivasan.

An estimator δ^* which is admissible and minimax for all possible full rank specifications of the design matrix when σ^2 is known to be one has been given by Berger [1]. Its form is reasonably simple: Let n be an integer such that $K/2 - 1 < n < K - 2$

and define $u = b'(X'X)\; Q(X'X)\; b/t_s$, where t_s is the smallest characteristic root of $(X'X)^{-1}Q$. Then an admissible minimax estimator is

$$\delta^*(b) = \left(I_K - (X'X)2nt_s^{-1}u^{-1}f(u)\right) b \;, \tag{3.22}$$

where $f(u) = 1 - (u/2)^n/n! \left(\exp(u/2) - \sum_{j=o}^{n} \frac{(u/2)^j}{j!}\right)$.

Other minimax admissible estimators [4] found in the literature have been restricted to certain kinds of design matrices. In particular, a minimax admissible estimator for $X'X = I_K$ and σ^2 not assumed to be known is given by Strawderman [11], but the form is complicated.

4. SUMMARY

In this paper we have tried to shed some light on the correction factor (shrink) type estimators that arise due to positing various forms of uncertainty in the general linear statistical model.

Estimators which are of the form $(I - H)b$ may be examined for possible improvement by taking the "positive part" of the estimator. Further, when $H = h(b'B\, b/s)C$, there are rules for determining whether or not the estimator improves upon the usual least squares estimator b, i.e., is minimax. Also an estimator is given which is admissible and improves on b for any full rank specification of the design matrix X.

References

[1] Berger, James O. "Admissible Minimax Estimation of a Multivariate Normal Mean with Arbitrary Quadratic Loss." The Annals of Statistics (1976) Volume 4, pp. 223-226.

[2] Berger, J. and Bock, M. E. "Eliminating Singularities of Stein-type Estimators of Location Vectors." Journal of the Royal Statistical Society, B, 38, pp. 166-170.

[3] Berger, James O. and Srinivasan, C. "Generalized Bayes Estimators in Multivariate Problems." Purdue University Statistics Department Technical Report #481, March 1977.

[4] Bock, M. E. "Minimax Estimators of the Mean of a Multivariate Normal Distribution." The Annals of Statistics (1976) Volume 3, pp. 209-218.

[5] Bock, M. E.; Yancey, T. A.; and Judge, G. G. "The Statistical Consequences of Preliminary Test Estimators in Regression." Journal of the American Statistical Association (1973), pp. 109-116.

[6] Casella, George "Minimax Ridge Regression Estimation". Unpublished Ph.D. Thesis, Statistics Department, Purdue University (1977).

[7] Dempster, A. P., Schatzoff, Martin, and Wermuth, Nanny. "A Simulation Study of Alternatives to Ordinary Least Squares." Journal of the American Statistical Association (1977) Volume 72, pp. 77-106.

[8] Efron, Bradley and Morris, Carl. "Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution." The Annals of Statistics (1976) Volume 4, pp. 11-21.

[9] James, W. and Stein, C. "Estimation with Quadratic Loss." Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (1960) pp. 361-379.

[10] Strawderman, W. E. "On the Existence of Proper Bayes Minimax Estimators of the Mean of a Multivariate Normal Distribution." Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (1970), pp. 51-55.

[11] Strawderman, W. E. "Proper Bayes Minimax Estimators of the Multivariate Normal Mean Vector for the Case of Common Unknown Variances." The Annals of Statistics (1973), Volume 1, pp. 1189-1194.

[12] Stein, C. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability (1950) pp. 197-206.