Statistical Papers 32, 155–165 (1991)

Statistical Papers
Statistische Hefte

Admissible linear estimators of the multivariate normal mean without extra information

L. Wang

Received: July 30, 1990

Suppose y is normally distributed with mean $\mu \in IR^n$ and covariance σ^2V, where $\sigma^2 > 0$ and $V > 0$ is known. The n. s. conditions that a linear estimator $Ay + a$ of μ be admissible in the class of all estimators of μ which depend only on y are derived. In particular, the usual estimator $\delta_0(y) = y$ is admissible in this class. The results are applied to the normal linear model and the admissibilities of many well-known linear estimators are demonstrated.

Keywords: admissibility, multivariate normal mean, unknown covariance, quadratic loss, linear estimation, Stein-type estimators, normal linear model.

1. Introduction

Assume the n-dimensional (n arbitrary) random vector y has a normal distribution $N(\mu, \sigma^2V)$, where V is a known positive definite matrix whereas $(\mu, \sigma^2) \in \wp = IR^n \times (0, \infty)$ are unknown parameters. Consider the problem of estimating μ by d under the quadratic loss

$$L(d, \mu, \sigma^2) = (d - \mu)' Q (d - \mu) \tag{1.1}$$

where Q is positive definite. Then the estimators $\delta(y)$ of μ are evaluated by the risk

$$R(\delta, \mu, \sigma^2) = E_{(\mu, \sigma^2)} L(\delta(y), \mu, \sigma^2) \tag{1.2}$$

If δ_1 and δ_2 are two estimators of μ, then δ_1 is said to be as good as δ_2, if

$$R(\delta_1, \mu, \sigma^2) \le R(\delta_2, \mu, \sigma^2)$$

for all $(\mu, \sigma^2) \in \wp$; and δ_1 is said to be better than δ_2, if in addition the inequality holds for at least one $(\mu_0, \sigma_0^2) \in \wp$. If $\mathfrak{S}$ is a class of certain estimators of μ, then an estimator $\delta(y)$ is said to be admissible for μ in $\mathfrak{S}$, if $\delta(y) \in \mathfrak{S}$ and $\mathfrak{S}$ contains no estimator which is better than $\delta(y)$.

A fundamental result of admissibility of a (homogeneous) linear estimator in the class of all estimators of μ was derived by Cohen (1966), for the case where the covariance $\sigma^2 V$ is known to be the indentity matrix I (and under the loss (1.1) with $Q = I$). For the more general case where $V = I$ and σ^2 is unknown, as remarked in that paper, the similar results may be derived under some additional conditions as given in James and Stein (1961). In particular, the inadmissibility of the estimator $\delta_0(y) = y$ (when $n \geq 3$) was established by producing a dominating Stein-type procedure, $\delta(y, s)$ say, in which an extra random variable s which is independent of y has to be observed. See also James and Stein (1961) and Gleser (1986). However due to a result of Cheng (1982), if such kind of extra information is not available [e. g. in the case of only one observation] or if only the estimatiors which depend only on y are considered, then $\delta_0(y) = y$ is admissible. Thus a characterization of the admissible linear estimators in this class is of theoretical interest and practical importance and this is the main aim of the present paper.

Throughout the paper we consider the admissibility in the class $\mathscr{A}$ of all estimators of the form $\delta(y)$, where δ are functions on $\mathbb{R}^n$ (and hence depend only on y). Thus when we say an estimator $\delta(y)$ is admissible, we mean that $\delta(y)$ is admissible in $\mathscr{A}$. In section 2 the necessary and sufficient conditions that a linear estimator be admissible are derived. It is shown that the Cohen (1966)'s result is essentially true with only one exception of $\delta_0(y) = y$. In section 3 the results obtained are applid to the normal linear model and the admissibility of many well-known (homogeneous) linear estimators is demonstrated.

It should be mentioned that the problem of admissibility in the class of linear estimators has already been completely solved. The main results may be found in, e. g. Rao (1976), LaMotte (1982), Mathew, Rao and Sinha (1984), Baksalary and Markiewicz (1988) and Klonecki and Zontek (1988). Note that in this case the distributional assumption is not necessary as the risk (1.2) depends on the first and second moments only.

If A is a matrix, then A', r(A) and $\Re(A)$ will stand for the transpose, the rank and the column space of A respectively; if A and B are two symmetric matrices, then $A < B$ ($A \leq B$) will mean that $B - A$ is positive (nonnegative) definite.

2. Admissible linear estimators in $\mathcal{A}$

In this section we charaterize the set of all linear estimators of μ in $\mathcal{A}$. It is well- known that if $\delta(y)$ is admissible for μ under the loss (1.1) for some $Q_0 > 0$, then $\delta(y)$ is admissible for μ under (1.1) for any $Q \geq 0$. (Shinozak 1975, Rao 1976, Lemma 3.1). Thus in the following we prove the results under (1.1) for $Q = I$. First we state the following interesting result which is a direct consequence of Cheng (1982, Theorem 1.2).

Lemma 1: If y has distribution $N(\mu, \sigma^2 I)$, then $\delta_0(y) = y$ is admissible for μ.

Proof: If $\delta_0(y)$ were not admissible, then there existed an estimator $\delta(y) \in \mathcal{A}$, such that

$$R(\delta, \mu, \sigma^2) \leq R(\delta_0, \mu, \sigma^2)$$

for all $(\mu, \sigma^2) \in \wp$ with inequality for at least one (μ_0, σ_0^2). Clearly $\delta(y) \neq \delta_0(y)$ over a set of positive Lebesque measure. Then there exist $\upsilon \in \mathrm{IR}^n$, $d > 0$ and $\varepsilon > 0$, such that $x \in \mathrm{IR}^n$ and $\|x - \upsilon\| \leq d$ imply $(\delta(x) - x)'(\delta(x) - x) \geq \varepsilon$, where $\|x\| = \max_{1 \leq j \leq n} |x_j|$. Hence for any $0 < \sigma^2 < 1$,

$$E_{(\upsilon, \sigma^2)}(\delta(y) - y)'(\delta(y) - y) \geq \frac{\varepsilon}{(2\pi\sigma^2)^{n/2}} \int_{\|x - \upsilon\| \leq d\sigma} \exp\left[-\frac{(x-\upsilon)'(x-\upsilon)}{2\sigma^2}\right] dx$$

$$\geq \frac{\varepsilon (2d)^n}{(2\pi)^{n/2}} \min_{\|x\| \leq d} \exp(-x'x/2) > 0$$

On the other hand,

$$\begin{aligned} E_{(\upsilon,\sigma^2)}(\delta(y) - y)'(\delta(y) - y) &\leq 2R(\delta, \upsilon, \sigma^2) + 2R(\delta_0, \upsilon, \sigma^2) \\ &\leq 4n\sigma^2 \quad \to 0, \text{ as } \sigma^2 \to 0 \end{aligned}$$

This is a contradiction. □

Next theorem shows that, except for $\delta_0(y) = y$, Cohen (1966)'s conditions are necessary and sufficient for admissibility. By $\mathcal{L}$ we denote the subset of $\mathcal{A}$ which contains all linear estimators $Ay + a$ of μ.

Theorem 1: If y has distribution $N(\mu, \sigma^2 I)$ and $A \neq I$, then the necessary and sufficient conditions that Ay be admissible for μ are that A is symmetric, all eigenvalues of A are between 0 and 1, and at most two of them equal 1.

Proof: Suppose Ay is inadmissible. Then there exists an estimator $\delta(y) \in \mathcal{A}$, such that

$$E(\delta(y) - \mu)'(\delta(y) - \mu) \leq E(Ay - \mu)'(Ay - \mu) \tag{2.1}$$

for all $(\mu, \sigma^2) \in \wp$ and the inequality holds for some (μ_0, σ_0^2). Consider a random vector z which has distribution N (υ, I). Then from (2.1).

$$E\left(\frac{1}{\sigma_0}\delta(\sigma_0 z) - \upsilon\right)'\left(\frac{1}{\sigma_0}\delta(\sigma_0 z) - \upsilon\right) \leq E(Az - \upsilon)'(Az - \upsilon)$$

for all $\upsilon \in \mathbb{R}^n$ and the inequality holds for $\upsilon_0 = \mu_0/\sigma_0$. Thus by Cohen (1966, Theorm 2.1), A cannot satisfy all conditions of the theorem.

Conversley, if Ay is admissible, then it is admissible in $\mathcal{L}$. By Rao (1976, Theorem 3.3), A is symmetric and has all eigenvalues between 0 and 1. Thus we need only to show that A has at most two eigenvalues which equal 1. To this end let P be the orthogonal matrix such that

$$P'AP = \mathrm{diag}(d_1, \ldots, d_n) = D$$

where $1 \geq d_1 \geq \ldots \geq d_n \geq 0$. Since Py has distribution $N(P\mu, \sigma^2 I)$, APy is admissible for $P\mu$ and hence $P'APy = Dy$ is admissible for $P'P\mu = \mu$. Now we show that if $d_1 = d_2 = \ldots = d_r = 1$ for $r \geq 3$, then we may construct an estimator in $\mathcal{A}$ which is better than Dy.

As $A \neq I$, $r \leq n - 1$. Let z and υ be the vectors of the first r elements of y and μ respectively, and define

$$\begin{aligned} s &= \sum_{j=r+1}^{n} \left(y_j - \frac{1}{n-r}\sum_{i=r+1}^{n} y_i\right)^2, & &\text{if } r \leq n-2 \\ &= y_n^2, & &\text{if } r = n-1 \end{aligned}$$

Then z and s are independent and have distributions N $(\upsilon, \sigma^2 I)$ and $\sigma^2\chi^2_{n-r}$ respectively. Define

$$\delta(y) = (u', d_{r+1}y_{r+1}, \ldots, d_n y_n)'$$

where

$$u = \left(1 - \frac{(r-2)s}{(n-r+2)z'z}\right)z$$

Then we have

$$\begin{aligned} &E(\delta(y) - \mu)'(\delta(y) - \mu) - E(Ay - \mu)'(Ay - \mu) \\ &= E(u - \upsilon)'(u - \upsilon) - r\sigma^2 < 0 \end{aligned}$$

for all $\upsilon \in \mathbb{R}^r$ and $\sigma^2 > 0$, because u is a well-known Stein-type estimator for υ, see James and Stein (1961). □

Now we consider the more general case N $(\mu, \sigma^2 V)$ with $V > 0$ and the linear estimators $Ay + a$. First we need the following lemma.

Lemma 2: $Ay + a$ is admissible for μ if and only if $a \in \mathfrak{R}(A - I)$ and Ay is admissible for μ.

Proof: If $Ay + a$ is admissible. Then it is admissible in $\mathcal{L}$. The condition $a \in \mathfrak{R}(A - I)$ follows from Rao (1976, Corrolary 3.2). Let $a = Ab - b$, $b \in \mathbb{R}^n$, then $Ay + a = A(y + b) - b$ and

$$E(Ay + a - \mu)'(Ay + a - \mu) = E(A(y + b) - (\mu + b))'(A(y + b) - (\mu + b))$$

That Ay is admissible for μ if and only if $A(y + b) - b$ is admissible for μ is easily seen from the fact that the parameter space $\wp$, the estimator space $\mathcal{A}$ and the relation "better than" among $\mathcal{A}$ are invariant under the transformations $T : \mu \to \mu + b$, $\delta(y) \to \delta(y + b) - b$. □

Theorem 2: If y has distribution N $(\mu, \sigma^2 V)$ with $V > 0$ known, then the necessary and sufficient conditions that $Ay + a$ be admissible for μ are

(i) $a \in \mathfrak{R}(A - I)$

(ii) $AV = VA'$

(iii) $AVA' \leq AV$

(iv) $r(A - I) \geq n - 2$ or $A = I$.

Proof: By Lemma 2 the n. s. conditions are (i) and that Ay is admissible for μ. The later is in turn equivalent to that $V^{-1/2}AV^{1/2}(V^{-1/2}y)$ is admissible for $V^{-1/2}\mu$. (Rao 1976, Theorem 3.1) [Note that in that theorem the result (b) is true if S has full row rank; and (c) is true if S has full column rank.] Now $V^{-1/2}y$ has distribu- tion N $(V^{-1/2}\mu, \sigma^2 I)$, by Lemma 1 and Theorem 1, it is necessary and sufficient that either $V^{-1/2}AV^{1/2} = I$, which is equivalent to $A = I$; or $V^{-1/2}AV^{1/2} = V^{1/2}A'V^{-1/2}$, which is equivalent to (ii), $(V^{-1/2}AV^{1/2})(V^{-1/2}AV^{1/2})' \leq V^{-1/2}AV^{1/2}$, which is equivalent to (iii), and $V^{-1/2}AV^{1/2}$

has at most two eigenvalues which equal 1, which is equivalent to $n - 2 \leq r(V^{-1/2}AV^{1/2} - I) = r(A - I)$. □

3. Admissibility in the normal linear model

In this section we consider the linear model in which y is normally distributed with mean $Ey = X\beta$ and covariance $\sigma^2 V$, where $X \in \mathbb{R}^{n \times p}$ and $V > 0$ are known matrices, whereas $\beta \in \mathbb{R}^p$ and $\sigma^2 > 0$ are unknown parameters. It is also assumed that $r(X) = p \leq n$.

Note that the estimators of β are different from that of μ in last section in the sense that they are now the mappings from $\mathbb{R}^n$ to $\mathbb{R}^p$ rather than from $\mathbb{R}^n$ to $\mathbb{R}^n$. But we will still use the notions $\mathcal{A}$ and $\mathcal{L}$ to denote the corresponding classes of estimators. These can be easily distinguished from the context. Let the usual least squares estimator be denoted by $b_{LS} = (X'V^{-1}X)^{-1}X'V^{-1}y$.

Theorem 3: Let $A \in \mathbb{R}^{p \times n}$ and $a \in \mathbb{R}^p$. Then $Ay + a$ is admissible for β if and only if

(i) $a \in \Re(AX - I)$

(ii) $XAV = VA'X'$

(iii) $XAVA'X' \leq XAV$

(iv) $r(AX - I) \geq p - 2$ or $AX = I$ with $n = p$

Proof: If $Ay + a$ is admissible, then it is admissible in $\mathcal{L}$ and (i) - (iii) follow from Baksalary and Markiewicz (1988, Corrolary 4). To show (iv), we observe by (ii),

$$A = (X'V^{-1}X)^{-1}X'A'X'V^{-1}$$
$$= AX(X'V^{-1}X)^{-1}X'V^{-1}$$

and hence $Ay + a = AXb_{LS} + a$ is admissible for β in the class of all estimators of β which are functions of b_{LS}. Note that b_{LS} has distribution $N(\beta, \sigma^2(X'V^{-1}X)^{-1})$.

By Theorem 2 it is necessary that either $r(AX - I) \geq p - 2$ or $AX = I$. The later together with (i) imply $Ay + a = b_{LS}$. It is well-known that b_{LS} is admissible only if $n = p$, as otherwise the residual $s = (y - Xb_{LS})'(y - Xb_{LS})$ is available to construct a Stein-type estimator in $\mathcal{A}$ which is better than b_{LS} (Berger 1976, Judge and Bock 1978, Chapter

10).

Conversely, the sufficiency of the case $AX = I$ and $n = p$ follows from Theorem 2 and the fact that $z = Ay = X^{-1}y$ has distribution $N(\beta, \sigma^2 X^{-1}V(X^{-1})')$ and the estimator class $\mathcal{A}$ is invariant under the transformation T: $\delta(y) \rightarrow \delta(X^{-1}y)$. For the other case, note that AX and XA have the same nonzero eigenvalues and hence, $r(AX - I) \geq p - 2$ is equivalent to $r(XA - I) \geq n - 2$. Then XA satisfies the conditions of Theorem 2 and hence XAy is admissible for $X\beta$, which implies that Ay is admissible for β (Rao 1976,Theorm 3.1). [See the note in the proof of Theorem 2.] By a similar argument of Lemma 2, (i) implies that Ay + a is admissible for β. $\square$

Next theorem gives the results of admissibility of the estimators of $X\beta$.

Theorem 4: Let $A \in IR^{n \times n}$, $a \in IR^n$. Then $Ay + a$ is admissible for $X\beta$ if and only if

(i) $a \in \Re(AX - X)$

(ii) $\Re(VA') \subset \Re(X)$

(iii) $AV = VA'$

(iv) $AVA' \leq AV$

(v) $r(A - I) \geq n - 2$ or $A = I$

Proof: Again the necessity of (i) - (iv) follows from Baksalary and Markiewicz (1988, Corrolary 3). To show (v), let (by ii and iii) $AV = VA' = XB$, $B \in IR^{p \times n}$. Then $A = XBV^{-1}$ and $Ay = XBV^{-1}y$ is admissible for $X\beta$, which implies that $BV^{-1}y$ is admissible for β. By Theorem 3, it is necessary that either $r(BV^{-1}X - I) \geq p - 2$ or $BV^{-1}X = I$ with $n = p$, which is easily seen to be equivalent to (v).

The sufficiency of (i) - (v) follows immediately from Theorem 2 and a similar argument of Lemma 2. $\square$

In the remaining part of this paper we apply the previous results to demonstrate the admissibility of some well-known (homogeneous) linear estimators. These estimators have the following general form:

$$b(A) = UAU'b_{LS} = UAD^{-1}U'X'V^{-1}y$$

where $A = \mathrm{diag}(a_1, \ldots, a_p)$, $0 \le a_j \le 1$, $j = 1, \ldots, p$, and U is the orthogonal matrix such that

$$U'X'V^{-1}XU = \mathrm{diag}(d_1, \ldots, d_p) = D$$

and $d_1 \ge d_2 \ge \ldots \ge d_p > 0$. Let us see some examples:

(1) The ridge estimator (Hoerl and Kennard 1970):

$$\begin{aligned} b_R(K) &= (X'V^{-1}X + UKU')^{-1}X'V^{-1}y \\ &= b\,(D(D+K)^{-1}) \end{aligned}$$

where $K = \mathrm{diag}(k_1, \ldots, k_p)$, $k_j \ge 0$, $j = 1, \ldots, p$ and at least one $k_j > 0$.

(2) The principal components estimator (Kendall 1957, Johnson, Reimer and Rothrock 1972):

$$\begin{aligned} b_{PC}(q) &= b_{LS} - (X'V^{-1}X)^{-1}U_q[U_q'(X'V^{-1}X)^{-1}U_q]^{-1}U'_q\, b_{LS} \\ &= b\,(\begin{pmatrix} I_q & 0 \\ 0 & 0 \end{pmatrix}) \end{aligned}$$

where $0 < q < p$ and U_q is the matrix of last $p - q$ columns of U.

(3) Linear Bayes estimator (Rao 1976):

$$\begin{aligned} b_B(W) &= UWU'X'(V + XUWU'X')^{-1}y \\ &= b(W(D^{-1} + W)^{-1}) \end{aligned}$$

where $W \ge 0$.

(4) The iteration estimator (Trenkler 1978):

$$\begin{aligned} b_I(m, \tau) &= \tau \sum_{j=0}^{m} (I - \tau X'V^{-1}X)^j X'V^{-1}y \\ &= b\,(I - (I - \tau D)^{m+1}) \end{aligned}$$

where $0 < \tau < d_1^{-1}$ and $m = 0, 1, \ldots$.

(5) The shrinkage least squares estimator (Mayer and Wilke 1973):

$$b_S(\alpha) = \alpha\, b_{LS} = b(\alpha I)$$

where $0 < \alpha < 1$.

Now we examine the conditons in Theorem 3 for any $b(A) = UAD^{-1}U'X'V^{-1}y$. It is easily seen that (i) - (iii) are always satisfied by the definition of A whereas (iv) becomes either $r\,(UAD^{-1}U'X'V^{-1}X - I) = r\,(AD^{-1}U'X'V^{-1}XU - I) = r\,(A - I) \geq p - 2$ or $A = I$ with $n = p$. Further these conditions ars easily examined for all estimators (1) - (5). Thus we have the following results.

Corrolary 1: Let $A = \text{diag}\,(a_1, \ldots, a_p)$, $0 \leq a_j \leq 1$, $j = 1, \ldots, p$.

(1) $UAUb_{LS}$ is admissible for β if and only if $r\,(A - I) \geq p - 2$ or $A = I$ with $n = p$.

(2) b_{LS} is admissible for β if and only if $p \leq 2$ or $n = p$.

(3) $b_R(K)$ is admissible for β if and only if $r\,(K) \geq p - 2$.

(4) $b_{PC}(q)$ is admissible for β if and only if $q \leq 2$.

(5) All $b_B(W)$, $b_I(m, \tau)$ and $b_S(\alpha)$ are admissible for β.

Acknowledgement

This paper was written when I was at the Fachbereich Statistik, Universität Dortmund. I would like to thank Professor Walter Krämer and Professor Götz Trenkler for many helpful discussions and comments. Support by Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

References

Baksalary, J. K. and Markiewicz, A. (1988). Admissible Linear Estimators In The General Gauss-Markov Model. J. Statist. Plann. Inference, 19, 349-359.

Berger, J. (1982). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. Ann. Statist., 4, 223-226.

Cheng, Ping (1982). Admissibility of simulaneous estimation of several parameters. J. Sys. Sci. & Math. Scis., 2(3), 176-195.

Cohen, A. (1966). All admissible linear estimates of the mean vector. Ann. Math. Statist., 37, 458-463.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased estimation for nonorthogonal problems. Technometrics, 12, 55-67.

James, W. and Stein, C. (1961). Estimation with quadratic loss. Proc. Fouth Berkeley Symp. Math. Statist., Prob. 1, 361-379.

Johnson, S. R. Reimer, S. C. and Rothrock, T. P. (1972). Principal Components and the Problem of multicollinarity. Metroeconomica, 25, 306-314.

Judge, G. G. and Bock, M. E. (1978). The Statistical Implication of Pre-Test and Stein-Rule Estimators in Econometrics. North-Holland, Asterdam.

Kendall, M. G. (1957). A course in multivariate analysis. Charles Griffin, London.

Klonecki, W. and Zontek, S. (1988). On the stucture of admissible linear estimators. J. Multivariate Anal., 24, 11-30.

LaMotte, L. R. (1982). Admissibility in linear estimation. Ann. Math. Statist., 10, 245-255.

Mathew, T., Rao, C. R. and Sinha, B. K. (1984). Admissible Linear Estimation In Singular Linear Models. Commun. Statist. - Theor. Meth., 13(24), 3033-3045.

Mayer, L. S. and Willke, T. A. (1973). On biased estimation in linear models. Technometrics, 15, 495-508.

Rao, C. R. (1976). Estimation of parameters in a linear model. Ann. Math. Statist., 4, 1023-1037.

Shinozaki, N. (1975). A study of generalized inverse of matrix and estimation with quadratic loss. Ph. D thesis, Keio University, Japan.

Trenkler, G. (1978). An iteration estimator for a linear model. COMPSTAT 1978: Wien, 125-131, Physica-Verlag, Wien.

Dr. Liqun Wang

Universität Basel

Institut für Statistik und Oekonometrie

Petersgraben 51, CH-4051 Basel

Switzerland

List of symbols which are not typewritten

1. $\mathscr{A}$: the class of all estimators of the form $\delta(y)$, where δ are functions on $\mathbb{R}^n$.
2. $\mathscr{L}$: the subclass of $\mathscr{A}$ which contains all linear estimators.