AD-A093 824 PITTSBURGH UNIV PA INST FOR STATISTICS AND APPLICATIONS F/G 12/1
SOME COMMENTS ON THE MINIMUM MEAN SQUARE ERROR AS A CRITERION O--ETC(U)
OCT 80 C R RAO F49620-79-C-0161
UNCLASSIFIED TR-80-21 AFOSR-TR-80-1343 NL

1 OF 1
AD A093824

END
DATE
FILMED
2-81
DTIC

SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED

LEVEL II

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

(3)

1. REPORT NUMBER: AFOSR-TR-80-1343	2. GOVT ACCESSION NO.: AD-A093 824
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): SOME COMMENTS ON THE MINIMUM MEAN SQUARE ERROR AS A CRITERION OF ESTIMATION.	5. TYPE OF REPORT & PERIOD COVERED: Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Professor C. Radhakrishna Rao	8. CONTRACT OR GRANT NUMBER(s): F49620-79-C-0161
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA. 15260	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, D. C. 20332	12. REPORT DATE: October 1980
	13. NUMBER OF PAGES: 39
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Based on a lecture delivered at the International Symposium on Statistics and Related Topics held in Ottawa (May 5-8, 1980

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is shown that estimators obtained by MMSE (minimizing the mean square error) may not have optimum properties with respect to other criteria such as PN (probability of nearness to the true value in the sense of Pitman) or PC (probability of concentration around the true value). In particular, a detailed study is made of estimators obtained by shrinking the minimum variance unbiased estimators to reduce the MSE. It is suggested that because of mathematical convenience and some intutive considerations, MMSE could be used as a primitive postulate to derive estimators, but their acceptabilty should be judged on

more intrinsic criteria such as PN and PC.

AFOSR-TR- 80-1343

SOME COMMENTS ON THE MINIMUM MEAN SQUARE

ERROR AS A CRITERION OF ESTIMATION*

C. Radhakrishna Rao

Accession For	
NTIS GRA&I	X
DTIC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

October 1980

Technical Report No. 80-21

Institute for Statistics and Applications

Department of Mathematics and Statistics

University of Pittsburgh

Pittsburgh, PA. 15260

The work is sponsored by the Air Force Office of Scientific Research under Contract [redacted]. F49620 79-C 0161
Reproduction in whole or in part is permitted for any purpose of the United States Government.

*Based on a lecture delivered at the International Symposium on Statistics and Related Topics held in Ottawa (May 5-8, 1980).

81 1 15 038

SOME COMMENTS ON THE MINIMUM MEAN SQUARE ERROR AS A CRITERION OF ESTIMATION

C. Radhakrishna Rao

Abstract. It is shown that estimators obtained by MMSE (minimizing the mean square error) may not have optimum properties with respect to other criteria such as PN (probability of nearness to the true value in the sense of Pitman) or PC (probability of concentration around the true value). In particular, a detailed study is made of estimators obtained by shrinking the minimum variance unbiased estimators to reduce the MSE. It is suggested that because of mathematical convenience and some intuitive considerations, MMSE could be used as a primitive postulate to derive estimators, but their acceptability should be judged on more intrinsic criteria such as PN and PC.

AMS(MOS) Subject Classification: 62F10, 62F15

Key Words and Phrases: Inverse regression, James-Stein estimator, Minimum mean square error, Shrunken estimator.

1. INTRODUCTION

The concept of minimum mean square error (MMSE) as a criterion of estimation is attributed to Gauss and figures prominently in the discussion of problems of statistical estimation. No doubt, the criterion is a valid one if the problem of estimation is considered in a decision theoretic frame work with the loss function specified as the square of the error in an estimator. Otherwise, the criterion is arbitrary as Gauss himself has observed in a paper presented to the Royal Society of Göttingen in 1809:

> "From the value of the integral $\int_{-\infty}^{\infty} x\phi(x)dx$, i.e., the average value of x (defined as deviation in the estimator from the true value of the parameter) we learn the existence or non-existence of a constant error as well as the value of this error; similarly, the integral $\int_{-\infty}^{\infty} x^2\phi(x)dx$, i.e., the average value of x^2, seems very suitable for defining and measuring, in a general way, the uncertainty of a system of observations. ... If one objects that this convention is arbitrary and does not appear necessary, we readily agree. The question which concerns us here has something vague about it from its very nature, and cannot be made really precise except by some principle which is arbitrary to a certain degree. ... It is clear to begin with that the loss should not be proportional to the error committed, for under this

hypothesis, since a positive error would be considered as a loss, a negative error would be considered as a gain; the magnitude of a loss ought, on the contrary, to be evaluated by a function of the error whose value is always positive. Among the infinite number of functions satisfying this condition, it seems natural to choose the simplest, which is, without doubt, the square of the error, and in this way we are led to the principle proposed above".

Karlin (1958) expresses the same opinion:

"The justification for the quadratic loss as a measure of the discrepancy of an estimate derives from the following two characteristics: (i) in the case where $a(x)$ represents an unbiased estimate of $h(\omega)$, MSE may be interpreted as the variance of $a(x)$ and, of course, fluctuations as measured by the variance is very traditional in the domain of classical estimation; (ii) from a technical and mathematical viewpoint square error lends itself most easily to manipulation and computations".

Thus, the criterion of MMSE is used not because of its practical relevance in a given problem but for its simplicity and mathematical convenience. We may, therefore, accept MMSE as a primitive postulate providing a rule of estimation like other methods such as maximum likelihood, minimum chi-square, etc., and examine the properties of estimators so obtained in terms of other criteria.

The present study is limited to the examination of estimators obtained by "shrinking" unbiased estimators with a view to decrease the MSE. We compare the shrunken estimator with the unbiased estimator in terms of its bias (B), mean absolute error (MAE), mean square error (MSE), mean quartic error (MQE), and more intrinsic properties like the probability of nearness to the true value (PN) due to Pitman (1937), and probability of concentration in intervals round the true value (PC).

In the discussion on a recent paper by Berkson (1980), the author (Rao, 1980) has pointed out some anamolies that may result in accepting MMSE as a criterion of estimation. Examples were given of estimators which have a smaller MSE but perform poorly in terms of more intrinsic criteria such as PN and PC when compared to other estimators. Such anamolies are expected since the quadratic loss function places undue emphasis on large deviations which may occur with small probability, and minimizing MMSE may insure against large errors in an estimator occurring more frequently rather than providing greater concentration of an estimator in neighborhoods of the true value. A more detailed study of such situations is made in the present paper.

2. ESTIMATION OF A SINGLE PARAMETER

Let X be an unbiased estimator of a parameter θ with $V(X) = \sigma^2$. It is well known that with respect to a quadratic loss function, cX is an admissible estimator of θ if $0<c\leq 1$ (see Rao, 1976b for instance). The MSE of cX is

$$E(cX-\theta)^2 = \sigma^2[c^2+(1-c)^2\delta^2] \leq E(X-\theta)^2 \qquad (2.1)$$

iff $\delta^2 \leq (1+c)/(1-c)$ where $\delta = \theta/\sigma$. Thus, if we have some knowledge of δ, we can make an appropriate choice of c to ensure the inequality in (2.1). The minimum of $E(cX-\theta)^2$ is attained at $c = \delta^2/(1+\delta^2)$, and if it is known that the true δ is near about δ_o, we may try the estimator

$$X_o = \frac{\delta_o^2}{1+\delta_o^2} X \qquad (2.2)$$

which has the property

$$E_2 = [E(X_o-\theta)^2/E(X-\theta)^2]^{\frac{1}{2}} \leq 1 \text{ if } |\delta| \leq (2\delta_o^2+1)^{\frac{1}{2}}. \qquad (2.3)$$

But the property (2.3) does not ensure that

$$PN = Pr.(|X_o-\theta| < |X-\theta|) \geq 0.5 \qquad (2.4)$$

for the same range of δ. Table 1 gives the approximate values of δ below which $PN \geq 0.5$ and $E_2 \leq 1$ for different values of the shrinkage factor $c = \delta_o^2/(1+\delta_o^2)$ and the associated values of δ_o.

TABLE 1

Values of $|\delta|$ below which $E_2 \leq 1$ and PN $\geq .05$

for different shrinkage factors

c	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
δ_o	0	.33	.50	.65	.82	1.00	1.22	1.53	2.00	3.00
$E_2 \leq 1$	1.0	1.2	1.3	1.4	1.6	1.8	2.0	2.5	3.0	4.5
PN $\geq$ 0.5	0.7	0.8	0.8	0.9	1.0	1.1	1.2	1.4	1.6	1.8

TABLE 1 shows that the range of δ for which (2.4) holds is much smaller than that for (2.3) to hold. It is also interesting to note that the optimum choice of c corresponding to a given δ_o for reducing the MSE does not ensure that PN $\geq$ 0.5 even for $\delta = \delta_o$ unless δ_o is below 1.2 (approximately). Thus, shrinking an unbiased estimator is useful only when the true value of the parameter under estimation is smaller than about 1.2 times the standard error of estimation.

If σ^2, the variance of the estimator X, is unknown, but an estimator s^2 of σ^2 is available, we can define an empirical version of (2.2)

$$X_e = \frac{(X/s)^2}{1+(X/s)^2} X \qquad (2.5)$$

and study its performance. The MSE of X_e compared to that of X has been extensively studied by Thompson (1968) under various distributional assumptions on X. We shall examine other properties of (2.5) assuming that X is normally distributed and

σ^2 is known. As shown by Thompson, the conclusions are not likely to be different when σ^2 is used instead of s^2 in (2.5) even for small values of f, the degrees of freedom on which σ^2 is estimated.

Table 2 gives the values of

$$B = \sigma^{-1} E(X_e - \theta) \quad , \quad PN = Pr.(|X_e - \theta| \leq |X - \theta|),$$

$$E_1 = \sigma^{-1}\sqrt{\pi/2}\ E|X_e - \theta|,\ E_2 = \sigma^{-1}[E(X_e - \theta)^2]^{\frac{1}{2}},\ E_4 = \sigma^{-1}[E(X_e - \theta)^4/3]^{\frac{1}{4}},$$

obtained by simulation. It is seen that the empirically shrunken estimator X_e is better than the unbiased estimator X only when $\delta \leq 1.4$ (approximately), i.e., when the standard error of the estimator of a parameter is more than 70% of the value of the parameter. But a serious drawback of the estimator (2.5) may be the large negative bias it has unless δ is very small or very large.

TABLE 2

Values of E_1, E_2, E_3, E_4, $E = \Sigma E_i$, PN and B for the estimator $X^3/(s^2+X^2)$ for different values of $\delta = \theta/\sigma$.

δ	B	E_1	E_2	E_4	E	PN
0.0	-.005	.549	.702	.813	.688	1.000
0.5	-.171	.757	.764	.818	.780	.706
1.0	-.290	.946	.893	.856	.898	.565
1.4	-.344	1.047	.993	.919	.986	.504
1.5	-.352	1.066	1.015	.938	1.006	.487
2.0	-.367	1.124	1.097	1.031	1.084	.444
2.5	-.350	1.136	1.131	1.091	1.120	.436
3.0	-.317	1.124	1.133	1.121	1.126	.437
3.5	-.283	1.105	1.118	1.118	1.114	.444
4.0	-.252	1.086	1.100	1.102	1.096	.453
8.0	-.130	1.026	1.037	1.035	1.033	.476
10.0	-.105	1.018	1.028	1.026	1.024	.480
20.0	-.055	1.007	1.017	1.015	1.013	.490
100.0	-.015	1.003	1.013	1.012	1.009	.495

3. ESTIMATION OF VARIANCE

If S^2 denotes the corrected sum of squares of n i.i.d observations from $N(\mu,\sigma^2)$, it is well known that $s^2=S^2/(n-1)$ is the minimum variance unbiased estimator of σ^2. But $s_2^2=S^2/(n+1)$ has smaller MSE than s^2 uniformly for all σ^2 and all n, so that s^2 is inadmissible as an estimator of σ^2 with respect to the MSE criterion. How does s_2^2 compare with s^2 with respect to other criteria? Table 3 gives the values of the following for different degrees of freedom (n-1):

$$B = E(s_2^2 - \sigma^2)/\sigma^2,$$

$$PN = Pr.(|s^2 - \sigma^2| \leq | \leq |s^2 - \sigma^2|),$$

$$PC = Pr.(-\log a \leq \log s^2 - \log \sigma^2 \leq \log a),$$

$$PC_2 = Pr.(-\log a \leq \log s_2^2 - \log \sigma^2 \leq \log a),$$

$$E_2 = [E(s_2^2 - \sigma^2)^2 / E(s^2 - \sigma^2)^2]^{\frac{1}{2}}.$$

It is seen that PN, the probability that s_1^2 is closer to σ^2 than s^2, is less than 0.5 uniformly for all σ^2 and for all n although E_2 is uniformly less than unity for all σ^2 and for all n. Similarly, log s^2 has a greater concentration probability in any symmetrical interval around log σ^2 than log s_2^2 uniformly for all σ^2 and all n. Thus shrinking the unbiased estimator s^2 has resulted in a smaller MSE but has not brought the estimator closer to the true value of σ^2 in any sense. The unbiased estimator s^2 seems to have better intrinsic properties than s_2^2.

It may be noted that the optimum shrinkage of s^2 depends on the loss function chosen. If instead of the MSE, we choose the MQE = $E(cs^2 - \sigma^2)^4$ as the loss, then the optimum c is a solution of the cubic equation

$$(1+\frac{6}{n-1})(1+\frac{4}{n-1})(1+\frac{2}{n-1})c^3 - 3(1+\frac{4}{n-1})(1+\frac{2}{n-1})c^2 + 3(1+\frac{2}{n-1})c - 1 = 0. \quad (3.1)$$

The estimator so obtained is denoted by s_4^2.

TABLE 3

Values of E_2, B, PN, PC and PC_2

for different degrees of freedom (DF)

D.F. (n-1)	E_2	B*	PN	PC (first row) and PC_2 (second row) a=1.5	a=2.0	a=2.5	a=3
1	.577	.677	.221	.193 .123	.322 .206	.413 .267	.480 .315
2	.707	.500	.264	.290 .213	.471 .349	.588 .442	.667 .511
3	.774	.400	.290	.360 .285	.571 .457	.695 .566	.772 .643
4	.816	.333	.308	.416 .345	.644 .540	.768 .658	.838 .734
5	.845	.286	.323	.463 .395	.701 .608	.821 .727	.883 .801
6	.866	.250	.334	.503 .440	.747 .663	.859 .780	.913 .849
7	.882	.222	.346	.539 .479	.784 .709	.888 .822	.935 .885
8	.894	.200	.352	.570 .514	.815 .747	.911 .855	.951 .911
9	.904	.182	.359	.599 .545	.840 .780	.928 .882	.963 .932
10	.912	.167	.365	.624 .574	.862 .808	.942 .903	.972 .947
20	.953	.091	.400	.793 .762	.963 .945	.992 .987	.998 .996
40	.976	.048	.428	.926 .912	.996 .994	.999+ .999+	.999+ .999+

*The shrinkage factor is (1-B) where B = Bias/σ^2

On the other hand, the optimum c which minimizes the $MAE = E|cs^2-\sigma^2|$ is the one which minimizes the function

$$(c-1) + 2\, G_{n-1}\left(\frac{n-1}{c}\right) - 2\, c\, G_{n+1}\left(\frac{n-1}{c}\right) \tag{3.2}$$

where

$$G_k(a) = \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} \int_0^a e^{-t/2}\, t^{\frac{k}{2}-1}\, dt.$$

The estimator so obtained may be denoted by s_1^2.

Table 4 gives the values of $E_i = [E(|s_i^2-\sigma^2|^i)/E(|s^2-\sigma^2|^i)]^{1/i}$, $B_i = E(s_i^2-\sigma^2)$, $PN = Pr.(|s_i^2-\sigma^2| \leq |s^2-\sigma^2|)$ and $PC - PC_i$ where

$$PC = Pr.(-\log a \leq \log s^2 - \log \sigma^2 \leq \log a)$$

$$PC_i = Pr.(-\log a \leq \log s_i^2 - \log \sigma^2 \leq \log a)$$

for i = 1 and 4. It is seen that s^2 performs better than s_1^2 and s^2 in terms of PN and PC. Among the estimators s_1^2, s_2^2 and s_4^2, s_1^2 appears to be better than s_2^2 and s_4^2. The results are not unexpected since the distribution of s^2 is skew on the right and minimization of an expression of the type $E(c\, s^2 - \sigma^2)^m$ pulls the estimator away from σ^2 in the region around and below the modal value of s^2.

It is not clear why in statistical literature much emphasis is laid on the estimation of σ^2 and not on σ although in practice the latter should be the parameter of direct interest. Unfortunately, none of the properties such as

TABLE 4

Values of B_i, E_i, PN_i and $PC - PC_i$ for $i = 1, 4$ for different degrees of freedom

	s_1^2							s_4^2						
				$PC - PC_1$							$PC - PC_4$			
DF	E_1	B_1^*	PN_1	a=1.5	a=2	a=2.5	a=3	E_4	B_4^*	PN_2	a=1.5	a=2	a=2.5	a=3
1	.78	.577	.24	.04	.07	.10	.11	.32	.764	.20	.11	.18	.22	.25
2	.85	.404	.29	.04	.07	.09	.10	.46	.619	.23	.14	.21	.24	.25
3	.89	.311	.31	.04	.07	.08	.08	.55	.521	.25	.14	.20	.22	.22
4	.91	.252	.33	.04	.06	.07	.07	.61	.450	.27	.14	.19	.20	.18
5	.93	.212	.35	.04	.05	.06	.05	.66	.396	.28	.13	.17	.17	.14
6	.94	.183	.36	.03	.05	.05	.04	.70	.354	.29	.13	.16	.14	.12
7	.94	.161	.37	.03	.04	.04	.03	.73	.320	.30	.13	.15	.12	.09
8	.96	.144	.38	.03	.04	.03	.02	.75	.292	.31	.12	.13	.10	.07
9	.96	.130	.38	.03	.04	.03	.02	.77	.268	.32	.11	.12	.09	.06
10	.96	.118	.39	.03	.04	.02	.01	.79	.248	.33	.11	.11	.07	.05
20	.98	.063	.42	.02	.03	.00	.00	.88	.142	.37	.07	.04	.01	.00
40	.99	.032	.44	.01	.00	.00	.00	.93	.077	.40	.03	.00	.00	.00

*The optimum shrinkage factor is 1-B where B = Bias/σ^2

unbiasedness and MMSE are preserved under transformations of estimators and parameters. For instance, the minimum variance unbiased estimator of σ is

$$s^* = \left(\frac{n-1}{2}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} s = ts \qquad (3.3)$$

which is different from s while the MMSE of σ is

$$s_2^* = \left(\frac{n-1}{2}\right)^{-\frac{1}{2}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} s \qquad (3.4)$$

which is different from s_2. Now

$$E(s^* - \sigma)^2 = \sigma^2(t^2 - 1) > 2\sigma^2(1 - \frac{1}{t}) = E(s - \sigma)^2 \qquad (3.5)$$

so that s has a smaller MSE than s^* as an estimation of σ.

We shall compare the relative performances of s and s^* as estimators of σ and of s^2 and $(s^*)^2$ as estimators of σ^2. Table 5 gives the values of the following for different degrees of freedom:

$$E_2 = [E(s-\sigma)^2/E(s^*-\sigma)^2]^{\frac{1}{2}},$$

$$PN_1 = Pr.(|s^*-\sigma| \leq |s-\sigma|),$$

$$PN_2 = Pr.(|(s^*)^2-\sigma^2| \leq |s^2-\sigma^2|),$$

$$PC = Pr.(-\log a \leq \log s^2 - \log \sigma^2 \leq \log a),$$

$$PC_2 = Pr.(-\log a \leq \log(s^*)^2 - \log \sigma^2 \leq \log a).$$

It is seen that although $E_2 \leq 1$ uniformly for all σ and DF so that s has a smaller MSE than s* as an estimator of σ, PN_1 is uniformly above 0.5 so that s* is nearer to σ more often than s. What is more interesting is that PN_2 is also uniformly above 0.5 indicating that $(s^*)^2$ is nearer to σ^2 more often than s^2. Further, $\log(s^*)^2$ has greater concentration around $\log \sigma^2$ than $\log s^2$ around $\log \sigma^2$ if the DF is not small and the interval chosen is not short. It appears that the biased estimator $(s^*)^2$ of σ^2 has better properties in terms of PN and PC than s^2, although highly inadmissible with respect to MSE.

4. DIRECT OR INVERSE REGRESSION

Consider a pair of random variables (θ, Y) such that

$$Y = \theta + \epsilon, \quad E(\epsilon)=0, \quad \mathrm{cov}(\theta,\epsilon)=0, \quad V(\epsilon)=\sigma_0^2. \tag{4.1}$$

In practice θ stands for the true value of a quantity (such as the cholesterol level of a blood sample) and Y is a measurement of θ subject to error. Only Y is observable and not θ, in which case the problem is one of estimating or predicting θ given Y.

From (4.1), the regression of Y on θ is θ itself so that the <u>inverse regression</u> estimate of θ is Y which is also an unbiased estimator of θ. On the other hand, if the mean (μ) and variance (σ_θ^2) of the unconditional distribution of θ is known, then the regression of θ on Y is

TABLE 5

Values of E_2, PN_1, PN_2, PC and PC_2 for various degrees of freedom

D.F.	E_2	PN_1	PN_2	PC (a = 1.5)	PC_2 (a = 1.5)	PC (a = 2)	PC_2 (a = 2)	PC (a = 2.5)	PC_2 (a = 2.5)	PC (a = 3.0)	PC_2 (a = 3.0)
1	.841	.625	.622	.193	.186	.322	.313	.413	.407	.480	.478
2	.912	.586	.585	.290	.284	.471	.467	.588	.590	.667	.675
3	.940	.570	.569	.360	.356	.571	.570	.695	.702	.772	.784
4	.954	.560	.559	.416	.413	.644	.646	.768	.776	.838	.850
5	.963	.553	.553	.463	.461	.701	.705	.821	.829	.883	.893
6	.969	.549	.548	.503	.502	.747	.751	.859	.868	.913	.923
7	.973	.545	.545	.539	.538	.784	.789	.888	.896	.935	.943
8	.976	.542	.542	.570	.570	.815	.820	.911	.918	.951	.957
9	.979	.539	.539	.599	.599	.840	.845	.928	.934	.963	.968
10	.982	.537	.537	.624	.623	.862	.867	.942	.947	.972	.976
20	.991	.526	.526	.793	.794	.963	.966	.992	.993	.998	.998
40	.995	.519	.519	.926	.927	.996	.997	.999+	.999+	.999+	.999+

$$\hat{\theta} = \mu + \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_o^2} (Y - \mu) \tag{4.2}$$

which provides a direct regression estimate of θ. In practice, the estimation procedure (4.2) can be implemented by estimating μ, σ_θ^2 and σ_o^2 from past data on Y (cholesterol determinations) on a large number of individuals (see Rao, 1973, p. 337), and updating the estimates as more data accumulate. The estimator $\hat{\theta}$ can be identified as the Bayes estimator using a quadratic loss function and a relevant prior distribution for θ.

Suppose that an individual's blood sample has been referred to a clinic for the determination of cholesterol and the clinic reports the measurement as Y. What should we record as the estimate of blood cholesterol for the individual, the unbiased estimator Y or the Bayes estimator $\hat{\theta}$ of (4.2) using a relevant estimated prior distribution? There has been considerable controversy on this subject, in a slightly different context, in the calibration problem (see Berkson, 1969; Halperin, 1970; Krutchkoff, 1967, 1969, 1971 and Williams, 1969). We shall examine this problem in the set up of (4.1) assuming that the parameters μ, σ_θ^2 of the prior distribution and the variance σ_o^2 of the error of measurement are known. Now

$$E(\hat{\theta}-\theta)^2 = \frac{\sigma_\theta^2 \sigma_o^2}{\sigma_\theta^2+\sigma_o^2} \le \sigma_o^2 = E(Y-\theta)^2 \tag{4.3}$$

and the strict inequality holds if $\sigma_o \neq 0$, so that the mean square error of prediction is smaller for $\hat{\theta}$. Does this mean that $\hat{\theta}$ is closer to θ than Y in some sense? To examine this question we have to consider the distributions of Y and $\hat{\theta}$ for given θ.

The MSE's and Y and $\hat{\theta}$ for given θ are

$$E[(Y-\theta)^2|\theta] = \sigma_o^2 \tag{4.4}$$

$$E[(\hat{\theta}-\theta)^2|\theta] = \sigma_o^2\delta^2(\delta^2+\lambda^2)/(1+\delta^2)^2 \tag{4.5}$$

where $(\theta-\mu)/\sigma_\theta=\lambda$ and $\delta=\sigma_\theta/\sigma_o$. From (4.4) and (4.5),

$$E[(\hat{\theta}-\theta)^2|\theta] \leq E[(Y-\theta)^2|\theta] \tag{4.6}$$

iff $\lambda^2 \leq (1+2\delta^2)/\delta^2$. Then the efficiency of $\hat{\theta}$ compared to Y with respect to MSE depends on the magnitude of the deviation of the true value of θ from the apriori mean. If the deviation is large, $\hat{\theta}$ is less efficient than Y.

The estimator Y is unbiased while the bias in $\hat{\theta}$ is

$$E[(\hat{\theta}-\theta)|\theta] = -\lambda\sigma_\theta\sigma_o^2/(\sigma_o^2+\sigma_\theta^2) \tag{4.7}$$

so that large values of θ are under-estimated and small values are over-estimated.

Table 6 gives for different combinations of δ and λ the values of

$$E_2 = [E\{(\hat{\theta}-\theta)^2|\theta\}/E(Y-\theta)^2]^{\frac{1}{2}},$$

$$PN = Pr.(|\hat{\theta}-\theta| < |Y-\theta|),$$

where the regions for which (i) $E_2 < 1$, $PN > 0.5$, (ii) $E_2 < 1$, $PN < 0.5$ and (iii) $E_n > 1$, $PN < 0.5$ are marked. It is seen that $\hat{\theta}$ performs better than Y when the error of measurement is large and the true value is near the mean of the apriori distribution. But if precise estimation of large deviations from the apriori mean is more important (as it should be in a problem like the estimation of blood cholesterol), Y should be preferred to $\hat{\theta}$.

TABLE 6

Values of E_2 (first entry) and PN (second entry) for different combinations of λ and δ

δ \ λ	0.5	1.0	1.5	2.0	2.5	3.0
0.5	.283 .835	.447 .679	.632 .535	.825 .411	1.020 .308	1.217 .226
1.0	.559 .742	.707 .528	.901 .375	1.118 .275	1.346 .209	1.581 .160
1.5	.729 .673	.832 .460	.979 .353	1.154 .294	1.346 .248	1.548 .207
2.0	.825 .615	.894 .435	1.000 .371	1.131 .329	1.281 .290	1.442 .253
2.5	.879 .569	.928 .433	1.005 .391	1.104 .356	1.219 .322	1.346 .290
3.0	.912 .535	.949 .439	1.006 .407	1.082 .376	1.171 .346	1.273 .318
5.0	.966 .487	.980 .461	1.004 .442	1.035 .422	1.075 .403	1.121 .384
10.0	.991 .490	.995 .480	1.001 .470	1.010 .461	1.020 .451	1.033 .441
15.0	.996 .493	.997 .487	1.000 .480	1.004 .474	1.009 .467	1.015 .460

5. SIMULTANEOUS ESTIMATION OF TWO PARAMETERS

Let $X_1 \sim N(\theta_1, \sigma^2)$, $X_2 \sim N(\theta_2, \sigma^2)$ and $fs^2 \sim \sigma^2 \chi^2(f)$ be independent random variables, and consider the following estimators of θ_1, θ_2

$$t_1 = \frac{X_1+X_2}{2} + c\,\frac{X_1-X_2}{2} \tag{5.1}$$

$$t_2 = \frac{X_1+X_2}{2} + c\,\frac{X_2-X_1}{2} \tag{5.2}$$

as alternatives to the unbiased estimators X_1 and X_2. Then

$$E(t_i-\theta_i)^2 = \sigma^2\left[\frac{1+c^2}{2} + \frac{(1-c)^2\delta^2}{4}\right], i = 1,2. \tag{5.3}$$

and the expected compound quadratic loss (ECQL) is

$$E\sum_1^2 (t_i-\theta_i)^2 = \sigma^2\left[1+c^2 + \frac{(1-c)^2\delta^2}{2}\right] \tag{5.4}$$

where $\delta = (\theta_1-\theta_2)/\sigma$. The expression (5.4) attains the minimum when $c = \delta^2/(2+\delta^2)$. Since δ^2 is not known, we may consider the empirical versions of (5.1) and (5.2)

$$t_1^{(e)} = \frac{X_1+X_2}{2} + \frac{(X_1-X_2)^2/s^2}{2+(X_1-X_2)^2/s^2}\,\frac{X_1-X_2}{2}$$

$$t_2^{(e)} = \frac{X_1+X_2}{2} + \frac{(X_1-X_2)^2/s^2}{2+(X_1-X_2)^2/s^2}\,\frac{X_2-X_1}{2}$$

We shall compare $t_1^{(e)}$ and $t_2^{(e)}$ with X_1 and X_2, assuming that   is known, with respect to the following criteria:

$$B_1 = \sigma^{-1}(t_1^{(e)}-\theta_1), \ B_2 = \sigma^{-1}(t_2^{(e)}-\theta_2),$$

$$PN = \frac{1}{2}\left[Pr.(|t_1^{(e)}-\theta_1| \leq |X_1-\theta_1|) + Pr.(|t_2^{(e)}-\theta_2| \leq |X_2-\theta_2|)\right.$$

$$E = (E_1 + E_2 + E_4)/3,$$

where

$$E_1 = \sigma^{-1}\sqrt{\pi/2}(\tfrac{1}{2}E|t_1^{(e)}-\theta_1| + \tfrac{1}{2}E|t_2^{(e)}-\theta_2|),$$

$$E_2 = \sigma^{-1}\left[\tfrac{1}{2}E(t_1^{(e)}-\theta_1)^2+\tfrac{1}{2}E(t_2^{(e)}-\theta_2)^2\right]^{\frac{1}{2}},$$

$$E_4 = \sigma^{-1}\left[\tfrac{1}{6}E(t_1^{(e)}-\theta_1)^4+\tfrac{1}{6}E(t_2^{(e)}-\theta_2)^4\right]^{\frac{1}{4}}.$$

Table 7 gives the values of E_1, E_2, E_4, E, PN and B_1, B_2 based on a simulation study using 1000 samples, for various values of $\delta = (\theta_1-\theta_2)/\sigma$. It is seen that simultaneous estimation of θ_1, θ_2 by $t_1^{(e)}$ and $t_2^{(e)}$ has some advantage over X_1 and X_2 when $\delta \leq 2$ (approximately), i.e., when the parameters under estimation do not differ by more than twice the standard error of the estimator of a single parameter.

6. ESTIMATION OF SEVERAL PARAMETERS

Let $X_i \sim N(\theta_i, \sigma^2)$, $i = 1, \ldots, p$ and $fs^2 \sim \sigma^2 \chi^2(f)$ be independent random variables, where $(\theta_1, \ldots, \theta_p) = \theta'$ is a

TABLE 7

Values of E_1, E_2, E_4, E, B_1, B_2 and PN for $t_1^{(e)}$ and $t_2^{(e)}$ for different values of $\delta = (\theta_1-\theta_2)/\sigma$

δ	B_1	B_2	PN	E_1	E_2	E_4	E
0	-.027	.005	.702	.854	.864	.872	.863
.5	.052	-.132	.674	.871	.871	.878	.873
1.0	.121	-.164	.635	.894	.887	.883	.888
1.5	.216	-.145	.567	.962	.956	.954	.957
2.0	.257	-.259	.517	1.013	.998	.976	.996
2.5	.269	-.277	.485	1.041	1.029	1.000	1.023
3.0	.199	-.243	.475	1.046	1.045	1.041	1.044
3.5	.279	-.234	.455	1.081	1.064	1.037	1.061
4.0	.229	-.251	.442	1.087	1.083	1.080	1.083
5.0	.188	-.132	.454	1.037	1.034	1.023	1.031
6.0	.207	-.181	.465	1.027	1.029	1.025	1.027
7.0	.197	-.118	.468	1.042	1.040	1.049	1.044
8.0	.085	-.102	.491	1.022	1.018	1.010	1.017
9.0	.070	-.108	.490	1.015	1.024	1.029	1.023
10.0	.097	-.105	.485	1.006	1.009	1.025	1.013

<u>fixed</u> vector parameter. James and Stein (1961) have found the remarkable result that when $p \geq 3$ there exist statistics

$$T_i = T_i(X_1, \ldots, X_p, s^2), \quad i = 1, \ldots, p \tag{6.1}$$

such that

$$E[\Sigma(T_i-\theta_i)^2] < E[\Sigma(X_i-\theta_i)^2] \tag{6.2}$$

uniformly for all θ_i, which implies that $X' = (X_1, \ldots, X_p)$ as an estimator of θ is inadmissible with respect to the CQL (compound quadratic loss) function. The result (6.2) gives the impression that we stand to gain by answering several problems, possibly unrelated, simultaneously. It is well known that there do not exist statistics t_i alternative to X_i such that

$$E(t_i-\theta_i)^2 < E(X_i-\theta_i)^2, \quad i=1,\ldots,p \tag{6.3}$$

uniformly for all θ_i, so that the overall reduction in the ECQL possibly takes place by an increase in the MSE for some parameters and decrease to a larger extent for the others. To examine this phenomenon in some detail, we shall consider a number of alternative joint estimators of $\theta_1,\ldots,\theta_p$ of the type suggested by James and Stein and study the performance of individual estimators.

Specifically, we consider the following types of estimators of $\theta_1,\ldots,\theta_p$:

$$T_{1i} = b\,X_i\,, \quad i=1,\ldots,p\,, \tag{6.4}$$

$$T_{2i} = a+b(X_i-a), \quad i=1,\ldots,p\,, \tag{6.5}$$

$$T_{3i} = a+b_i(X_i-a), \quad i=1,\ldots,p\,, \tag{6.6}$$

which may be represented by T_1, T_2 and T_3 in vector notation.

Now

$$E[\Sigma(T_{1i}-\theta_i)^2] = p\,b^2\,\sigma^2 + (1-b)^2\Sigma\theta_i^2 \tag{6.7}$$

which attains the minimum value at $b = \nu^2/(1+\nu^2)$ where $\nu^2 = \Sigma\theta_i^2/p\sigma^2$. If ν is known, then the optimum estimator of the type (6.4) is

$$T_1 = \left(1 - \frac{1}{1+\nu^2}\right) \underline{X} \tag{6.8}$$

and the ECQL is

$$E[\Sigma(T_{1i}-\theta_i)^2] = p\sigma^2 \frac{\nu^2}{1+\nu^2} \geq p\sigma^2 = E[\Sigma(X_i-\theta_i)^2]. \tag{6.9}$$

The MSE for an individual estimator is

$$E(T_{1i}-\theta_i)^2 = \sigma^2(\nu^4+\nu_i^2)/(1+\nu^2)^2 \tag{6.10}$$

where $\nu_i = \theta_i/\sigma$. The expression exceeds the MSE of X_i if $\nu_i^2 > 2\nu^2+1$ indicating the possibility that in joint estimation of the T_1-type, the larger parameters are less efficiently estimated than the smaller ones.

If ν^2 is not known, we may estimate $1/(1+\nu^2)$ by $c\,s^2/\Sigma X_i^2$, where c is a suitable constant, and obtain an empirical version of (6.8)

$$T_1^{(e)} = \left(1- \frac{c\,s^2}{\Sigma X_i^2}\right) X. \tag{6.11}$$

The best choice of c obtained by minimizing the ECQL of

(6.11) is $f(p-2)/(f+2)$ if $p \geq 3$, which leads to the James-Stein estimator

$$T_1^{(e)} = \left(1 - \frac{f(p-2)}{f+2} \frac{s^2}{\Sigma X_i^2}\right) X. \tag{6.12}$$

James and Stein have shown that for $p \geq 3$

$$E[(T_1^{(e)}-\theta)'(T_1^{(e)}-\theta)] = p\sigma^2 - \frac{f(p-2)^2\sigma^2}{f+2} E[(2K_1+p-2)^{-1}] \tag{6.13}$$

where K_1 is a Poisson variable with parameter $\nu^2/2$. The expression (6.13) is smaller than $p\,\sigma^2$ for all ν. Ullah (1974) computed the bias and MSE of individual estimators:

$$\sigma^{-1} E(T_{1i}^{(e)}-\theta_i) = -\frac{(p-2)f}{f+2} \nu_i E[(2K_1+p)^{-1}] \tag{6.14}$$

$$\sigma^{-2} E(T_{1i}^{(e)}-\theta_i)^2 = 1 - \frac{f(p-2)^2}{2(f+2)} \left[(1+g)E\{(2K_1+p-2)^{-1}\} - (1+\frac{pg}{p-2})E\{(2K_1+p)^{-1}\}\right] \tag{6.15}$$

where

$$g = [(p+2)\theta_i^2 - 2\,\Sigma\,\theta_i^2] \,/\, \Sigma\,\theta_i^2.$$

Similarly, the best estimator of the type (6.5) is

$$T_{2i} = \bar{\theta} + \left(1 - \frac{1}{1+\eta^2}\right)(X_i - \bar{\theta}), \quad i = 1, \ldots, p \tag{6.16}$$

where $\bar{\theta} = (\theta_1 + \ldots + \theta_p)/p$ and $\eta^2 = \Sigma(\theta_i - \bar{\theta})^2/p\,\sigma^2$. The ECQL for T_2 as in (6.16) is

$$p\,\sigma^2 \left(\frac{\eta^2}{1+\eta^2}\right) \leq p\,\sigma^2 \left(\frac{\nu^2}{1+\nu^2}\right) \leq p\,\sigma_o^2 \tag{6.17}$$

so that if $\overline{\theta}$ and η^2 are known further improvement in ECQL over X and T_1 is possible. The MSE for an individual estimator is

$$E(T_{2i}-\theta_i)^2 = \sigma^2(\eta^4+\eta_i^2)/(1+\eta^2)^2 \qquad (6.18)$$

where $\eta_i^2 = (\theta_i-\overline{\theta})^2/\sigma^2$. The expression (6.18) exceeds σ^2, the MSE of X_i, if $\eta_i^2 > 2\eta^2+1$, indicating the possibility that in joint estimation of the type T_2, the extreme parameters are less efficiently estimated and the middle parameters more efficiently than the corresponding unbiased estimators.

As in the case of T_1, we can estimate $\overline{\theta}$ and $1/(1+\eta^2)$ by $\overline{X}$ and $(p-3)fs^2/(f+2)[\Sigma(X_i-\overline{X})^2]$ respectively if $p \geq 4$ and obtain an empirical version of T_2

$$T_{2i}^{(e)} = \overline{X} + \left(1 - \frac{f(p-3)}{(f+2)} \frac{s^2}{\Sigma(X_i-\overline{X})^2}\right)(X_i-\overline{X}), \qquad (6.19)$$

$$i = 1,\ldots,p(\geq 4).$$

It can be shown (Efron and Morris, 1971, 1972a, Rao, 1976a) that

$$E\{(T_2^{(e)}-\theta)'(T_2^{(e)}-\theta)\} = p\,\sigma^2 - \frac{f(p-3)^2\sigma^2}{f+2} E\{(2K_2+p-3)^{-1}\} \quad (6.20)$$

where K_2 is a Poisson variable with parameter $\eta^2/2$. The expression (6.20) is less than $p\,\sigma_o^2$ so that $T_2^{(e)}$ is uniformly better than X with respect to ECQL. Rao and Shinozaki (1978) have shown that for individual estimators

$$E(T_{2i}-\theta_i) = \frac{-f(p-3)}{f+2}(\theta_i-\overline{\theta})E\{(2K_2+p-1)^{-1}\} \qquad (6.21)$$

$$E(T_{2i}-\theta)^2 = \sigma^2 - \frac{f(p-3)^2\sigma^2}{2(f+2)} \qquad (6.22)$$

$$\times\left[(c+d)E\{(2K_2+p-3)^{-1}\} - \frac{c(p-3)+d(p-1)}{(p-3)}E\{(2K_2+p-1)^{-1}\}\right]$$

where

$$c=(p-1)/p,\quad d=\{(p+1)(\theta_i-\overline{\theta})^2 - \frac{2(p-1)}{p}\Sigma(\theta_i-\overline{\theta})^2\}/\Sigma(\theta_i-\overline{\theta})^2.$$

Finally, the best estimator of the type (6.6) is

$$T_{3i} = \overline{\theta} + \frac{(\theta_i-\overline{\theta})^2}{(\theta_i-\overline{\theta})^2+\sigma_o^2}(X_i-\overline{\theta}),\quad i=1,\ldots,p \qquad (6.23)$$

and its empirical version is

$$T_{3i}^{(e)} = \overline{X} + \frac{(X_i-\overline{X})^2}{(X_i-\overline{X})^2+s^2}(X_i-\overline{X}),\quad i=1,\ldots,p\,. \qquad (6.24)$$

It is difficult to compute the ECQL of $\underline{T}_3^{(e)}$ or the MSE of $T_{3i}^{(e)}$.

The relative performance of the estimators $T_1^{(e)}$ and $T_2^{(e)}$ and the ranges of parametric values for which the individual $T_1^{(e)}$ and $T_2^{(e)}$- estimators are better than the X-estimators are examined in Rao and Shinozaki (1978). Table 8 contains the results of simulation studies based on thousand samples for the estimation of four parameters $\theta_1 = a\sigma$, $\theta_2=(a+d)\sigma$, $\theta_3=(a+2d)\sigma$ and $\theta_4=(a+3d)\sigma$ for various combinations of a and d, assuming σ to be known, i.e., by replacing $fs^2/(f+2)$ by σ^2 in the formulae (6.12) and (6.19) for $T_2^{(e)}$ and $T_3^{(e)}$. The broad conclusions remain the same if the random variable s^2 is used provided f, the degrees of freedom, is not small.

For each combination of $\bar{\theta}/\sigma$ and d, Table 8 gives the values of E, PN and B in the first row for $T_1^{(e)}$, in the second row for $T_2^{(e)}$ and in the third row for $T_3^{(e)}$, where for any statistic

$$E = [E(t_i-\theta_i)^2 \div E(X_i-\theta_i)^2]^{\frac{1}{2}},$$

$$PN = Pr.(|t_i-\theta_i| \le |X_i-\theta_i|),$$

$$B = E(t_i-\theta_i).$$

On the basis of previous investigations by Efron and Morris (1971, 1972a,b, 1973a,b), Rao (1975a, 1975b, 1975c, 1977) and Rao and Shinozaki (1978) and the present simulation studies, the following broad conclusions emerge.

(a) There is some advantage in using $T_2^{(e)}$ and $T_3^{(e)}$ when the range of parameter values is small, and $T_1^{(e)}$ when both the range and values of the parameters are small compared to the standard error of the unbiased estimators.

(b) When the range of parameters is large both $T_1^{(e)}$ and $T_2^{(e)}$ tend to have the same properties as X. But the performance of $T_3^{(e)}$ tends to be erratic.

TABLE 8

Values of E_2, PN and B for estimators of the type T_1, T_2 and T_3 for various combinations of $\bar{\theta}/\sigma$ and d

$\frac{\bar{\theta}}{\sigma}$	d		E_2				PN				Bias			
			θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
1.25	0.5	T_1	.835	.884	.955	1.069	.770	.589	.479	.440	.094	-.183	-.297	-.410
		T_2	.927	.863	.864	.950	.634	.766	.751	.620	.197	.079	-.067	-.209
		T_3	.761	.614	.613	.771	.692	.727	.723	.692	.246	-.092	.105	.232
2.00	1.0	T_1	.919	.947	.985	.953	.793	.521	.446	.453	.041	-.123	-.215	-.304
		T_2	1.004	.912	.914	1.010	.500	.739	.721	.510	.217	.078	-.074	-.220
		T_3	.814	.688	.693	.816	.652	.712	.704	.645	.028	-.015	.019	-.033
2.75	1.5	T_1	.950	.972	.993	1.030	.800	.478	.456	.466	-.022	-.088	-.157	-.230
		T_2	1.015	.949	.950	1.016	.476	.696	.662	.480	.175	.061	-.059	-.177
		T_3	.901	.793	.802	.899	.575	.626	.612	.570	.160	.005	-.006	-.159
3.50	2.0	T_1	.972	.983	.995	1.019	.912	.462	.465	.476	-.013	-.068	-.127	-.183
		T_2	1.011	.969	.970	1.012	.482	.646	.597	.482	.139	.048	-.047	-.140
		T_3	.967	.888	.894	.964	.521	.486	.477	.524	.224	.008	-.011	-.220
7.00	4.0	T_1	.993	.997	.999	1.004	.651	.464	.481	.488	-.057	-.036	-.065	-.093
		T_2	1.003	.991	.992	1.003	.501	.536	.500	.492	.073	.025	-.024	-.073
		T_3	1.054	1.035	1.035	1.056	.465	.199	.201	.469	.229	.000	-.001	-.226

(c) When the range of parameters is moderate, $T_1^{(e)}$ gives higher precision for the parameters with smaller values at the expense of lower precision for higher values. According to the PN criterion, only the smallest of the four parameters is better estimated than the corresponding unbiased estimator. In the case of $T_2^{(e)}$ and $T_3^{(e)}$, extreme values of the parameters suffer at the expense of increased precision for the middle values.

(d) One drawback in using estimators $T_1^{(e)}$, $T_2^{(e)}$ and $T_3^{(e)}$ in preference to X, is the bias in these estimators. The bias is of a substantial magnitude for the higher values in the case of $T_1^{(e)}$ and for the extreme values of the parameters in the case of $T_2^{(e)}$ and $T_3^{(e)}$. There are situations where bias in the estimators may have serious consequences such as the following.

Suppose there are four regions and periodical estimates of a certain characteristic are needed for sharing some resources in proportion to the values of the characteristic of the four regions. If each time, estimates of the type $T_1^{(e)}$, $T_2^{(e)}$ and $T_3^{(e)}$ are used, some regions stand to gain at the expense of the others in the long run (Rao, 1977, 1979).

In some situations, the individual parameters may not be of direct interest but certain linear combinations may be important. If $c'\theta = c_1\theta_1+\ldots+c_4\theta_4$ is a linear combination to be estimated, should one estimate it by $c'X$ or $c'T_1^{(e)}$ or $c'T_2^{(e)}$ or $c'T_3^{(e)}$? Naturally, the answer depends on the vector c (Rao, 1975b, 1975c), and the optimal properties of $T_1^{(e)}$, $T_2^{(e)}$ and $T_3^{(e)}$ with respect to a single criterion like the CQL do not insure their efficiency in different ways they may be used for practical purposes. If multiple uses are intended, the best plan is to place on record X as the estimator of θ (together with an estimate of σ^2) leaving it to the user to make any optimal adjustments in X depending on particular problems under study.

<u>Note 1</u>. It may be of historical interest to note that estimators of the type T_2 have been constructed under more general conditions, in multivariate analysis, for purposes of genetic selection by Fairfield Smith (1936), Hazel (1943) and Rao (1953) based on an idea suggested by Fisher. The problem was as follows. Let $(\underline{\theta}, \underline{y}_1, \ldots, \underline{y}_m)$ be (m+1) vector variables representing the unobservable genetic values $\underline{\theta}$ and repeated independent phenotypic vector measurements $\underline{y}_1, \ldots, \underline{y}_m$ on an individual. The variables are related by the model

$$\underline{y}_i = \underline{\theta} + \underline{\epsilon}_i \ , \ i = 1, \ldots, m. \tag{6.24}$$

The genetic worth of an individual is measured by a linear function $\underline{g}'\underline{\theta}$. Suppose that we have observed p individuals from a population, with phenotypic measurements

$$(\underline{y}_{1j}, \ldots, \underline{y}_{mj}), \ j = 1, \ldots, p. \tag{6.25}$$

What is the best way estimating the genetic worths $\underline{g}'\underline{\theta}_1, \ldots, \underline{g}'\underline{\theta}_p$ of these individuals for purposes of ranking and selecting a given proportion of the individuals with the largest genetic values? If $\overline{\underline{y}}_j$ represents the mean of the measurements (6.25) for individual j, then $\underline{g}'\overline{\underline{y}}_j$ is an unbiased estimator of $\underline{g}'\underline{\theta}_j$. However, Fisher suggested the regression of $\underline{g}'\underline{\theta}_j$ on $\overline{y}_j$ as the appropriate selection index, which involves the knowledge of $\underline{\mu} = E(\underline{\theta})$, $\Gamma = \mathrm{cov}(\theta,\theta)$, $\Delta = \mathrm{cov}(\overline{\epsilon},\overline{\epsilon})$, where $\overline{\underline{\epsilon}} = m^{-1}\,\Sigma\,\underline{\epsilon}_i$. The regression estimator of $\underline{g}'\underline{\theta}_j$ is $\underline{g}'\hat{\underline{\theta}}_j$ where

$$\hat{\theta}_j = \mu + [I - \Delta(\Gamma+\Delta)^{-1}](\overline{\underline{y}}_j - \underline{\mu}). \qquad (6.26)$$

By multivariate analysis of variance and covariance of the data (6.25), we obtain dispersion matrices B and W as between and within individuals with degrees of freedom (p-1) and $f = m(p-1)$ respectively, which supply estimates $B/(p-1)$ of $(\Gamma+\Delta)$ and W/mf of Δ. Then an empirical version of (6.26) is

$$\hat{\underline{\theta}}_j = \overline{\underline{y}} + [I - \frac{p-1}{mf} W B^{-1}] (\overline{\underline{y}}_j - \overline{\underline{y}}) \qquad (6.27)$$

where $p\overline{\underline{y}} = \Sigma\,\overline{y}_j$. The details leading to the formula (6.27) are given in Rao (1953, pp. 237-8). When all the variables are one dimensional, (6.27) is the same as $T_{2j}^{(e)}$ of (6.19) except that the multiplying factor $(p-1)/f$ is replaced by $(p-3)(f+2)$. In the 1953 paper, Rao also considered some distributional problems for testing hypotheses concerning the rank of the Γ matrix and the efficiency of the regression estimator.

It should be noted that the regression estimators (or empirical Bayes estimators) $\underline{g}'\hat{\underline{\theta}}_j$ are appropriate in the problem of selection where the total genetic worth of the selected subset of individuals has to be maximized. In such a case, it is well known that the best ordering of the observed individuals is achieved by using, as the selection index, the regression of genetic worth on phenotypic measurements (see Cochran, 1951 and Henderson, 1963). But the regression estimator may not be appropriate if the genetic worth of each individual has to be assessed for other purposes which may demand equal precision for the individual estimators.

Note 2. In his presidential address delivered to the Royal Statistical Society, Finney (1974) suggested that the problem of simultaneous estimation may be approached through the principle of maximum likelihood, thus avoiding the use of the arbitrary compound quadratic loss function. Let $X_i \sim N(\theta_i, \phi)$, $i = 1,\ldots,p$ be p independent observations. If θ_i arise as a random sample from $N(\mu, \tau)$, then the log likelihood is, apart from a constant,

$$L = -\frac{\Sigma(X_i-\theta_i)^2}{2\,\phi} - \frac{\Sigma(\theta_i-\mu)^2}{2\,\tau}. \qquad (6.28)$$

Finney maximizes L with respect to μ and θ_i and obtains the estimates

$$\hat{\theta}_i = \bar{X} + \left(1 - \frac{\phi}{\tau+\phi}\right)\left(X_i-\bar{X}\right). \qquad (6.29)$$

It is not known whether the maximum likelihood principle applies in situations such as (6.28) where the likelihood is a function of both the unknown parameters and unobserved random variables. Finney says that an unbiased estimate of $1/(\tau+\phi)$

is $(p-3)/\Sigma(X_i-\bar{X})^2$, so that when τ is not known, the estimator of θ_i is

$$\hat{\theta}_i = \bar{X} + \left(1 - \frac{(p-3)\phi}{\Sigma(X_i-\bar{X})^2}\right)(X_i-\bar{X}) \tag{6.30}$$

which is the same as the expression given by Lindley (1962) using Bayes theorem and quadratic loss function.

However, if τ is unknown, the appropriate log likelihood is proportional to

$$-\frac{p}{2}\log\tau - \frac{\Sigma(X_i-\theta_i)^2}{2\phi} - \frac{\Sigma(\theta_i-\mu)^2}{2\tau} . \tag{6.31}$$

The expression (6.31) can be made arbitrarily large by choosing $\theta_i = \mu = \bar{X}$ for all i and letting $\tau \to 0$. Thus, the m.l. estimator is $\theta_i = \bar{X}$ for all i! Such anamolies do occur when unobserved random variables are included as unknowns and a "full likelihood" function such as (6.31) is considered for drawing inference.

I wish to thank Robert Boudreau for the simulation studies reported in the various tables.

REFERENCES

Berkson, J. (1969). Estimation of a linear function for a calibration line; consideration of a recent proposal. Technometrics, 11, 649-660.

Berkson, J. (1980). Minimum chi-square, not maximum likelihood. Ann. Math. Statist., 8, 457-469.

Cochran, W. G. (1951). Improvement by means of selection. Proc. 2nd Berkeley Symp. on Math. Statist. Prob., 449-470.

Efron, B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators, I. The Bayes case. J. Amer. Statist. Assoc., 66, 807-815.

Efron, B. and Morris, C. (1972a). Empirical Bayes on vector observations. Biometrika 55, 335-347.

Efron, B. and Morris, C. (1972b). Limiting the risk of Bayes and empirical Bayes estimators, II. The empirical Bayes case. J. Amer. Statist. Assoc., 67, 130-139.

Efron, B. and Morris, C. (1973a). Stein's estimation rule and its competitors--an empirical Bayes approach. J. Amer. Statist. Assoc., 68, 117-130.

Efron, B. and Morris, C. (1973b). Combining possibly related estimation problems. J. Roy. Statist. Soc. Ser. B., 35, 379-421.

Fairfield Smith, H. (1936). A discrimant function for plant selection. Ann. Eugen., 7, 240-260. London.

Finney, D. J. (1974). Problems, data and inference. J. Roy. Statist. Soc., 137, 1-23.

Halperin, M. (1970). On inverse estimation in linear regression. Technometrics, 12, 727-736.

Hazel, L. N. (1943). The genetic basis for constructing selection indexes. Genetics, 28, 476.

Henderson, C. R. (1963). Selection index and expected genetic advance. Statistical Genetics and Plant Breeding NAS-NRC 982, 141-163.

James, W. and Stein, C. (1961). Estimation with quadratic loss. Proc. 4th Berkeley Symp. 1, 362-379.

Karlin, Samuel (1958). Admissibility for estimation with quadratic loss. Ann. Math. Statist., 29, 406-436.

Krutchkoff, R. G. (1967). Classical and inverse regression methods of calibration. Technometrics, 9, 425-439.

Krutchkoff, R. G. (1969). Classical and inverse regression methods of calibration in extrapolation. Technometrics, 11, 605-608.

Krutchkoff, R. G. (1971). The calibration problem and closeness. J. Statist. Comput. Simul., 1, 87-95.

Lindley, D. V. (1962). Contribution to discussion on paper by C. M. Stein. J. Roy. Statist. Soc. Ser. B., 24, 285-287.

Pitman, E. J. G. (1937). The closest estimates of statistical parameters. Proc. Camb. Philos. Soc., 32, 212.

Rao, C. R. (1953). Discriminant function for genetic differentiation and selection. Sankhyā 12, 229-246.

Rao, C. R. (1973). Linear Statistical Inference and its Applications. Wiley, New York.

Rao, C. R. (1975a). Simultaneous estimation of parameters in different linear models and applications to biometric problems. Biometrics 31, 545-554.

Rao, C. R. (1975b). Some thoughts on regression and prediction. Sankhyā C 37, 102-120.

Rao, C. R. (1975c). Some problems of sample surveys. Advances in Appl. Probability Supp., 7, 50-61.

Rao, C. R. (1976a). Characterization of prior distribution and solution to a compound decision problem. Ann. Statist. 4, 823-835.

Rao, C. R. (1976b). Estimation of parameters in a linear model-Wald Lecture 1. Ann. Math. Statist., 4, 1023-1037.

Rao, C. R. (1977). Simultaneous estimation of parameters--a compound decision problem. In Decision Theory and Related Topics, (S. S. Gupta and D. S. Moore, Eds.). p. 327-350. Academic Press, New York.

Rao, C. R. (1979). Presidential address-42nd Session of the International Statistical Institute, Manila.

Rao, C. R. (1980). Discussion on a paper by Berkson. Ann. Math. Statist., 8, 482-485.

Rao, C. R. and Shinozaki, N. (1978). Precision of individual estimators in simultaneous estimation of parameters. Biometrika, 65, 23-30.

Thompson, J. R. (1968). Some shrinkage techniques for estimating the mean. J. Amer. Statist. Assoc., 63, 113-122.

Ullah, A. (1974). On the sampling distribution of improved estimators for coefficients in linear regression. J. Econometrics, 2, 143-150

Williams, E. J. (1969). A note on regression methods in calibration. Technometrics, 11, 189-192.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER **AFOSR-TR- 80 - 1343**	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Some Comments on the Minimum Mean Square Error as a Criterion of Estimation		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) C. Radhakrishna Rao		8. CONTRACT OR GRANT NUMBER(s) F49620-79-0161
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Pittsburgh Department of Mathematics and Statistics Pittsburgh, PA. 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling Air Force Base Washington, D. C. 20332		12. REPORT DATE October 1980
		13. NUMBER OF PAGES 39
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Based on a lecture delivered at the International Symposium on Statistics and Related Topics held in Ottawa (May 5-8, 1980)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Inverse regression, James-Stein estimator, Minimum mean square error, Shrunken estimator

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is shown that estimators obtained by MMSE (minimizing the mean square error) may not have optimum properties with respect to other criteria such as PN (probability of nearness to the true value in the sense of Pitman) PC (probability of concentration around the true value). In particular, a detailed study is made of estimators obtained by shrinking the minimum variance unbiased estimators to reduce the MSE. It is suggested that because of mathematical convenience and some intuitive considerations, MMSE could

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) be used as a primitive postulate to derive estimators, but their acceptability should be judged on more intrinsic criteria such as PN and PC.