BIOMETRICS 31, 545–554
June 1975

SIMULTANEOUS ESTIMATION OF PARAMETERS IN DIFFERENT LINEAR MODELS AND APPLICATIONS TO BIOMETRIC PROBLEMS

C. RADHAKRISHNA RAO

Indian Statistical Institute, Planning Unit, New Delhi, India

SUMMARY

Empirical Bayes procedure is employed in simultaneous estimation of vector parameters from a number of Gauss-Markoff linear models. It is shown that with respect to quadratic loss function, empirical Bayes estimators are better than least squares estimators. While estimating the parameter for a particular linear model, a suggestion has been made for distinguishing between the loss due to decision maker and the loss due to individual. A method has been proposed but not fully studied to achieve balance between the two losses. Finally the problem of predicting future observations in a linear model has been considered.

1. INTRODUCTION

We consider k linear models

$$\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i\,, \; i = 1, \cdots, k, \tag{1.1}$$

where $\mathbf{Y}_i$ is an n-vector of observations, $\mathbf{X}$ is a known $n \times m$ matrix, and $\boldsymbol{\beta}_i$ is an m-vector and $\boldsymbol{\varepsilon}_i$ is an n-vector of unobservable random variables. We assume that

$$E(\boldsymbol{\varepsilon}_i \mid \boldsymbol{\beta}_i) = 0, \; D(\boldsymbol{\varepsilon}_i \mid \boldsymbol{\beta}_i) = \sigma^2\mathbf{V} \tag{1.2}$$

$$E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}, \; D(\boldsymbol{\beta}_i) = \mathbf{F}, \; \text{cov}\,(\boldsymbol{\beta}_i\,, \boldsymbol{\beta}_j) = 0, \; i \neq j. \tag{1.3}$$

(In (1.2) and (1.3), operator E stands for expectation and D for dispersion, i.e., variances and covariances.) The following problems will be considered.

1. Simultaneous estimation of $\mathbf{p}'\boldsymbol{\beta}_i\,, i = 1, \cdots, k$, where $\mathbf{p}$ is any given vector. We note that the problem of estimating $\boldsymbol{\beta}_i$ is the same as that of estimating a general linear function $\mathbf{p}'\boldsymbol{\beta}_i$. If we use the criterion of *minimum mean square error* (MSE) in estimating $\mathbf{p}'\boldsymbol{\beta}_i$, we automatically obtain estimate of $\boldsymbol{\beta}_i$ with a *minimum mean dispersion error matrix* (MDE).

Such a problem of simultaneous estimation arises in the construction of a selection index for choosing individuals with a high intrinsic genetic value. For instance, $\boldsymbol{\beta}_i$ may represent unknown genetic parameters and $\mathbf{Y}_i$ are observable characteristics on the ith individual, while $\mathbf{p}'\boldsymbol{\beta}_i$ for given $\mathbf{p}$ is the genetic value to be estimated in terms of observed $\mathbf{Y}_i$. Early examples of such estimation (which may be called empirical Bayes) by computing the regression of $\mathbf{p}'\boldsymbol{\beta}_i$ on $\mathbf{Y}_i$ (suggested by R. A. Fisher) is due to Fairfield Smith [1936] and Panse [1946]. A detailed study of this problem from a decision theoretic view point with an estimated prior distribution of $\boldsymbol{\beta}_i$ is given by Rao [1953]. Some applications are given in Rao [1952, 1953].

Interest in the problem of simultaneous estimation of parameters was revived by James and Stein [1961]. They showed that individual unbiased estimators of *unrelated* scalar parameters can be uniformly improved with respect to a quadratic loss function.

Research in this direction is being pursued by Effron [1974] and Effron and Morris [1972, 1973a, 1973b]. A slight modification of the James-Stein estimator is given in Rao [1974b] and the caution required in using such estimators is discussed in Rao [1974a].

Recently a number of authors considered this problem (see, e.g., Lindley and Smith [1972], Smith [1973] and Swamy [1970] and the references given in their papers). Lindley and Smith [1972] use Bayesian methods assuming suitable prior distributions. We shall review some of these methods in section 2, but our emphasis will be more on empirical Bayes procedures as discussed in Rao [1952, 1953, 1965, 1973] leading to the James-Stein type of estimators.

2. Suppose we have a $(k + 1)$th linear model in addition to the past k linear models (1.1),

$$\mathbf{Y}_{k+1} = \mathbf{X}\boldsymbol{\beta}_{k+1} + \boldsymbol{\varepsilon}_{k+1} \tag{1.4}$$

and the parameter to be estimated is only $\mathbf{p}'\boldsymbol{\beta}_{k+1}$, the parameters $\mathbf{p}'\boldsymbol{\beta}_1, \cdots, \mathbf{p}'\boldsymbol{\beta}_k$ being no longer of interest. Do the observations $\mathbf{Y}_1, \cdots, \mathbf{Y}_k$ in (1.1) obtained in the past contain information on the *current parameter* $\boldsymbol{\beta}_{k+1}$? If so, how can they be used in addition to $\mathbf{Y}_{k+1}$ for estimating $\mathbf{p}'\boldsymbol{\beta}_{k+1}$?

A problem of this type was mentioned and solved in a simple case by the author in recent papers (Rao[1974a, 1974b]). Individuals are continuously observed from a population and on each individual measurements are obtained, such as blood pressure which are subject to error. In such a case an observation y_i on the ith individual has the structure

$$y_i = \beta_i + \epsilon_i \tag{1.5}$$

where β_i is the true value and ϵ_i is the error such that $E(\epsilon_i \mid \beta_i) = 0$ and $V(\epsilon_i \mid \beta_i) = \sigma^2$ The object is to estimate β_i, the true value for the ith individual currently under observation. Estimation of true values for individuals observed in the past may not be of current interest, but the observations on these $(i - 1)$ individuals may be useful in estimating the current parameter β_i. The problem posed in (1.4) is an extension of the simpler problem considered in Rao [1974b].

3. Suppose that in the $(k + 1)$th model, only the first $(n - r)$ components of the vector $\mathbf{Y}_{k+1}$ have been observed. How can the last r component of $\mathbf{Y}_{k+1}$ be predicted on the basis of the past observations $\mathbf{Y}_1, \cdots, \mathbf{Y}_k$ and $(n - r)$ components of $\mathbf{Y}_{k+1}$ observed on the current $(k + 1)$th individual?

Such a problem arises in prediction of growth. For instance, the components of $\mathbf{Y}_i$ may represent the heights of the ith individual observed at n points of time and $\mathbf{Y}_i$ has the structure (1.1) with $\boldsymbol{\beta}_i$ as a parameter specific to the ith individual. We have records of observed heights at n points of time on each of k individuals and for the first $(n - r)$ points of time on a $(k + 1)$th individual. How can all the available data be used to predict the heights of the $(k + 1)$th individual at the last r time points.

The problem of prediction of growth has been extensively studied by Geisser [1970, 1971, 1974], Lee and Geisser [1972, 1973], Lee [1972] and others. These authors examine a wide variety of prediction procedures and compare their relative efficiencies. Of course, the success of any method depends on the validity of an assumed growth model. A broad conclusion that has emerged from various studies is that growth curves are highly individualistic in nature and data on completed growth curves of individuals observed in the past may not contribute very much to the efficiency of prediction of growth in future cases. This is somewhat disturbing. In the present paper we consider the problem of prediction under a general linear model and obtain some results to supplement those of Geisser and Lee.

We use the following notations and results throughout the paper.

Consider a linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1.6}$$

where $\boldsymbol{\beta}$ is an m-vector of unknown parameters and $E(\boldsymbol{\epsilon}) = 0$, $D(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V}$. To avoid some complications, let us assume that $\mathbf{V}$ is nonsingular and rank of $\mathbf{X}$ is m.

The least squares estimator of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^{(l)} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \tag{1.7}$$

and a ridge regression estimator of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^{(r)} = (\mathbf{G} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \tag{1.8}$$

form some chosen non-negative (positive or positive semi) definite matrix G. (Ridge regression estimator was introduced by Hoel and Kennard [1970a, 1970b] in the special case $\mathbf{V} = \mathbf{I}$ with the particular choice $\mathbf{G} = k^2\mathbf{I}$.) It may be noted that

$$\boldsymbol{\beta}^{(r)} = \mathbf{T}\boldsymbol{\beta}^{(l)} \tag{1.9}$$

where $\mathbf{T} = (\mathbf{G} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ has all its latent roots less than unity if $\mathbf{G}$ is not the null matrix. The following matrix identities, which are well known and quoted in Smith [1973] will prove useful:

$$(\mathbf{V} + \mathbf{X}\mathbf{F}\mathbf{X}')^{-1} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \mathbf{F}^{-1})^{-1}\mathbf{X}'\mathbf{V}^{-1} \tag{1.10}$$

$$(\mathbf{V} + \mathbf{X}\mathbf{F})^{-1} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{I} + \mathbf{F}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{F}\mathbf{V}^{-1} \tag{1.11}$$

$$(\mathbf{V} + \mathbf{F})^{-1} = \mathbf{V}^{-1} - \mathbf{V}^{-1}(\mathbf{V}^{-1} + \mathbf{F}^{-1})^{-1}\mathbf{V}^{-1} \tag{1.12}$$

(See also Rao [1973] p. 33).

We shall say that matrix $\mathbf{A}$ is larger than matrix $\mathbf{B}$ if $\mathbf{A} - \mathbf{B}$ is non-negative definite.

2. ESTIMATION OF PARAMETERS FROM DIFFERENT LINEAR MODELS

Let us consider k linear models

$$\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \; i = 1, \cdots, k, \tag{2.1}$$

as mentioned in (1.1) with

$$E(\boldsymbol{\epsilon}_i \mid \boldsymbol{\beta}_i) = 0, \qquad D(\boldsymbol{\epsilon}_i \mid \boldsymbol{\beta}_i) = \sigma^2\mathbf{V} \tag{2.2}$$

$$E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}, \qquad D(\boldsymbol{\beta}_i) = \mathbf{F} \tag{2.3}$$

where $\mathbf{X}$ and $\mathbf{V}$ are known matrices with full rank. We shall find a linear function $a_0 + \mathbf{a}_1'\mathbf{Y}_i$ such that

$$E(\mathbf{p}'\boldsymbol{\beta}_i - a_0 - \mathbf{a}_1'\mathbf{Y}_i)^2 \tag{2.4}$$

is a minimum for given $\mathbf{p}$. The problem as stated is easily solvable when σ^2, $\boldsymbol{\beta}$ and $\mathbf{F}$ are known (see e.g., Rao [1965, 1973] section 4a.11, Lindley and Smith [1972]). We shall review known results and also consider the problem of estimation when σ^2, $\boldsymbol{\beta}$ and $\mathbf{F}$ are unknown but can be estimated.

2.1 *Case 1* (σ^2, β *and* F *are known*)

Theorem 1: *The optimum estimator of* $p'\beta_i$ *in the sense of* (2.4) *is* $p'\beta_i^{(b)}$ *where* $\beta_i^{(b)}$ *can be written in the following alternative forms (where* $U = (X'V^{-1}X)^{-1}$)

$$\beta_i^{(b)} = \beta + FX'(XFX' + \sigma^2V)^{-1}(Y_i - X\beta) \tag{2.5}$$

$$= \beta + (\sigma^2F^{-1} + U^{-1})^{-1}X'V^{-1}(Y_i - X\beta) \tag{2.6}$$

$$= (\sigma^2F^{-1} + U^{-1})^{-1}\sigma^2F^{-1}\beta + \beta_i^{(r)} \tag{2.7}$$

$$= \beta + F(F + \sigma^2U)^{-1}(\beta_i^{(l)} - \beta) \tag{2.8}$$

$$= \sigma^2U(F + \sigma^2U)^{-1}\beta + F(F + \sigma^2U)^{-1}\beta_i^{(l)} \tag{2.9}$$

$$= \beta_i^{(l)} - \sigma^2U(F + \sigma^2U)^{-1}(\beta_i^{(l)} - \beta). \tag{2.10}$$

where $\beta_i^{(r)}$ *is the ridge regression estimator as defined in* (1.8) *with* $G = \sigma^2F^{-1}$. *The prediction error is* $p'Qp$ *where*

$$Q = \sigma^2(\sigma^2F^{-1} + U^{-1})^{-1} \tag{2.11}$$

$$= \sigma^2F(F + \sigma^2U)^{-1}U \tag{2.12}$$

$$= \sigma^2(U - \sigma^2U(F + \sigma^2U)^{-1}U). \tag{2.13}$$

Some of the results are proved in section 4a.11 of Rao [1965, 1973] and others can be easily deduced using the identities (1.10)–(1.12). We shall refer to $\beta_i^{(b)}$ as the Bayes estimator of β_i with parameters of its prior distribution as defined in (2.3). We make the following observations.

Note 1: Barnard [1974] has noted in a recent paper that the ridge regression estimator (1.8) originally defined with $V = I$ and $G = k^2I$ is Bayes estimator when the prior distribution of the regression parameter has 0 (null vector) as the mean and σ^2k^2I as the dispersion matrix. More generally, we find from (2.7) that the ridge regression estimator as defined in (1.8) is Bayes estimator when the mean and dispersion matrix of prior distribution are the null vector and σ^2G^{-1}, respectively.

Note 2: The Bayes estimator of β_i is a weighted linear combination of its least squares estimator and the mean of its prior distribution.

Note 3: The estimator $\beta_i^{(b)}$ as defined in Theorem 1 is optimum in the class of linear estimators. However, it is optimum in the entire class of estimators if the regression of β_i on Y_i is linear. A characterization of the prior distribution of β_i is obtained in Rao [1974b] using the property that the regression of β_i on Y_i is linear.

Note 4: The matrix

$$E(\beta_i^{(l)} - \beta_i)(\beta_i^{(l)} - \beta_i)' - E(\beta_i^{(b)} - \beta_i)(\beta_i^{(b)} - \beta_i)' \tag{2.14}$$

is non-negative definite. Of course, Bayes estimator has the minimum MDE compared to any other linear estimator.

Thus when σ^2, β and F are known, $p'\beta_i$ is estimated by $p'\beta_i^{(b)}$ for $i = 1, \cdots, k$ and the compound loss

$$E\sum^{K}(p'\beta_i - p'\beta_i^{(b)})^2 \tag{2.15}$$

is minimum compared to any other set of linear estimators. We shall consider the modifications to be made when σ^2, β, F are unknown.

Note 5: It may be noted that for fixed β_i, the expected value of (2.14) may not be non-negative definite. Indeed, as shown later in section 3, the optimality of Bayes estimator over the least squares estimator is not uniform for all values of β_i. It is true only for a region of β_i such that $||\beta_i - \beta||$, the norm of $\beta_i - \beta$ where β is the chosen prior mean of β_i, is less than a preassigned quantity depending on σ^2, $\mathbf{F}$ and $\mathbf{U}$.

2.2 *Case 2* σ^2, β, *and* $\mathbf{F}$ *are unknown*)

When σ^2, β and $\mathbf{F}$ are unknown, we shall substitute for them suitable estimates in the formulae (2.5)–(2.10) for estimating β_i. The following unbiased estimates σ_*^2, β_*, and $\mathbf{F}_*$ of σ^2, β and $\mathbf{F}$ are well known.

$$k\beta_* = \sum_1^k \beta_i^{(l)} \tag{2.16}$$

$$k(n - m)\sigma_*^2 = \sum_1^k (\mathbf{Y}_i'\mathbf{V}^{-1}\mathbf{Y}_i - \mathbf{Y}_i'\mathbf{V}^{-1}\mathbf{X}\beta_i^{(l)}) = \mathbf{W} \tag{2.17}$$

$$(k - 1)(\mathbf{F}_* + \sigma_*^2\mathbf{U}) = \sum_1^k (\beta_i^{(l)} - \beta_*)(\beta_i^{(l)} - \beta_*)' = \mathbf{B} \tag{2.18}$$

Substituting constant multiplies of these estimators for σ^2, β and $\mathbf{F}$ in (2.10) we obtain the empirical Bayes estimator of $\mathbf{p}'\beta_i$ as $\mathbf{p}'\beta_i^{(e)}$ where $\beta_i^{(e)}$ is

$$\beta_i^{(e)} = \beta_i^{(l)} - c\mathbf{W}\mathbf{U}\mathbf{B}^{-1}(\beta_i^{(l)} - \beta_*),\ i = 1, \cdots, k, \tag{2.19}$$

with $c = (k - m - 2)/(kn - km + 2)$ as determined in (2.30).

Theorem 2: *Let β_i and ε_i have multivariate normal distributions in which case W and* B *are independently distributed with*

$$\mathbf{W} \sim \sigma^2\chi^2(kn - km) \tag{2.20}$$

$$\mathbf{B} \sim W_m(k - 1, \mathbf{F} + \sigma^2\mathbf{U}) \tag{2.21}$$

i.e., as chi-square on $k(n - m)$ degrees of freedom and Wishart on $(k - 1)$ degrees of freedom, respectively. Then

$$E\sum_{i=1}^k (\beta_i^{(e)} - \beta_i)(\beta_i^{(e)} - \beta_i)' = k\sigma^2\mathbf{U} - \frac{\sigma^4k(n - m)(k - m - 2)}{k(n - m) + 2}\mathbf{U}(\mathbf{F} + \sigma^2\mathbf{U})^{-1}\mathbf{U} \tag{2.22}$$

for the optimum choice $c = (k - m - 2)/(kn - km + 2)$ in (2.19) *provided* $k \geq m + 2$.

Consider

$$\sum_1^k (\beta_i^{(e)} - \beta_i)(\beta_i^{(e)} - \beta_i)'$$

$$= \sum_1^k (\beta_i^{(l)} - \beta_i)(\beta_i^{(l)} - \beta_i)' + c^2\mathbf{W}^2\mathbf{U}\mathbf{B}^{-1}\mathbf{U} - 2c\mathbf{W}\mathbf{U}$$

$$+ c\mathbf{W}\sum_1^k \beta_i(\beta_i^{(l)} - \beta_*)'\mathbf{B}^{-1}\mathbf{U} + c\mathbf{W}\mathbf{U}\mathbf{B}^{-1}\sum_1^k (\beta_i^{(l)} - \beta_*)\beta_i'. \tag{2.23}$$

Let us observe that

$$E(\mathbf{W}) = k(n - m)\sigma^2, \qquad E(\mathbf{W}^2) = k(n - m)(kn - km + 2)\sigma^4 \tag{2.24}$$

$$E(\mathbf{B}^{-1}) = (k - m - 2)^{-1}(\mathbf{F} + \sigma^2\mathbf{U})^{-1} \tag{2.25}$$

$$E \sum_{1}^{k} \beta_i(\beta_i^{(1)} - \beta_*)'B^{-1} = F(F + \sigma^2 U)^{-1} \tag{2.26}$$

$$E\, B^{-1} \sum_{1}^{k} (\beta_i^{(1)} - \beta_*)\beta_i' = (F + \sigma^2 U)^{-1}F. \tag{2.27}$$

Then (2.23) reduces to

$$k\sigma^2 U + \sigma^4 g U(F + \sigma^2 U)^{-1} U \tag{2.28}$$

where

$$g = \frac{c^2 k(n - m)(kn - km + 2)}{k - m - 2} - 2ck(n - m). \tag{2.29}$$

The optimum choice of c in (2.29) is

$$c = (k - m - 2)/(kn - km + 2) \tag{2.30}$$

which leads to the value (2.22) given in Theorem 2.

Note 1: The results of Theorem 2 are generalizations of the results in the estimation of scalar parameters considered in an earlier paper of the author Rao [1974b].

Note 2: The expression (2.22) for the compound loss of empirical Bayes estimators is somewhat larger than the corresponding expression for Bayes estimators, which is k times (2.2), and the difference is the additional loss due to using estimates of σ^2, β and F when they are unknown.

Note 3: If β_i is estimated by $\beta_i^{(1)}$, then the compound MDE is

$$E \sum_{1}^{k} (\beta_i^{(1)} - \beta_i)(\beta_i^{(1)} - \beta_i)' = k\sigma^2 U \tag{2.31}$$

and the difference between (2.31) and (2.22), the MDE for empirical Bayes estimator, is

$$\frac{\sigma^4 k(n - m)(k - m - 2)}{k(n - m) + 2} U(F + \sigma^2 U)^{-1} U \tag{2.32}$$

which is non-negative definite.

Thus the expected compound loss for the estimation of $p'\beta_i$, $i = 1, \cdots, k$, is smaller for the empirical Bayes estimator than for the least squares estimator.

Note 4: It may be easily shown that the expectation of (2.23) for fixed values of $\beta_1, \cdots, \beta_k$ is smaller than $k\sigma^2 U$, as in the univariate case (Rao, 1974b). Thus the empirical Bayes estimators (2.19) are uniformly better than the least squares estimators without *any* assumption on the priori distribution of β_i. The actual expression for the expectation of (2.23) for fixed $\beta_1, \cdots, \beta_k$ has been worked out by C. G. Khatri in the form

$$k\sigma^2 U - \frac{\sigma^4 (k - m - 2)^2 k(n - m)}{k(n - m) + 2} E(UB^{-1}U)$$

which gives an indication of the actual decrease in loss by using empirical Bayes estimators.

Note 5: In the specification of the linear models we have assumed that the dispersion matrix $(\sigma^2 V)$ of the error vector is known apart from a constant multiplier. If V is unknown, it cannot be completely estimated from the observations $Y_1, \cdots, Y_k$ alone. However, if V has a suitable structure, it may be possible to estimate it. Such problems will be considered in a later paper.

2.3 *Estimation under different assumptions on linear models*:

Let us consider k linear models

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \; i = 1, \cdots, k, \tag{2.33}$$

where $\mathbf{X}_i = n_i \times m$ matrix and

$$E(\boldsymbol{\epsilon}_i \mid \boldsymbol{\beta}_i) = \mathbf{0}, \qquad D(\boldsymbol{\epsilon}_i \mid \boldsymbol{\beta}_i) = \sigma_i^2\mathbf{V}_i\,, \tag{2.34}$$

$$E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}, \qquad D(\boldsymbol{\beta}_i) = \mathbf{F}. \tag{2.35}$$

If σ_i^2, $\boldsymbol{\beta}$ and $\mathbf{F}$ are known, then the optimum estimate of $\boldsymbol{\beta}_i$ is, using formula (2.10),

$$\boldsymbol{\beta}_i^{(b)} = \boldsymbol{\beta}_i^{(l)} - \sigma_i^2\mathbf{U}_i(\mathbf{F} + \sigma_i^2\mathbf{U}_i)^{-1}(\boldsymbol{\beta}_i^{(l)} - \boldsymbol{\beta}) \tag{2.36}$$

where

$$\boldsymbol{\beta}_i^{(l)} = (\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{Y}_i\,,\; \mathbf{U}_i = (\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1} \tag{2.37}$$

The prediction error is

$$\sigma_i^2\mathbf{U}_i - \sigma_i^2\mathbf{U}_i(\mathbf{F} + \sigma^2\mathbf{U}_i)^{-1}\mathbf{U}_i\sigma_i^2 \tag{2.38}$$

If σ_i^2, $\boldsymbol{\beta}$ and $\mathbf{F}$ are unknown, we may use suitable estimates and substitute them in (2.36) to obtain empirical Bayes estimates,

$$\boldsymbol{\beta}_i^{(e)} = \boldsymbol{\beta}_i^{(l)} - \sigma_{i_*}^2\mathbf{U}_i(\mathbf{F}_* + \sigma_{i_*}^2\mathbf{U}_i)^{-1}(\boldsymbol{\beta}_i^{(l)} - \boldsymbol{\beta}_*). \tag{2.39}$$

The following estimates are suggested,

$$(n_i - m)\sigma_{i_*}^2 = \mathbf{Y}_i'\mathbf{V}_i^{-1}\mathbf{Y}_i - \mathbf{Y}_i'\mathbf{V}^{-1}\mathbf{X}_i\boldsymbol{\beta}_i^{(l)}, \; i = 1, \cdots, k \tag{2.40}$$

$$\boldsymbol{\beta}_* = k^{-1} \sum \boldsymbol{\beta}_i^{(l)} \tag{2.41}$$

$$(k-1)(\mathbf{F}_* + k^{-1} \sum \mathbf{U}_i\sigma_{i_*}^2) = \mathbf{B} = \sum (\boldsymbol{\beta}_i^{(l)} - \boldsymbol{\beta}_*)(\boldsymbol{\beta}_i^{(l)} - \boldsymbol{\beta}_*)' \tag{2.42}$$

$$(k-1)\mathbf{F}_* = \mathbf{B} - (k-1)k^{-1} \sum \mathbf{U}_i\hat{\sigma}_{i_*}^2 \tag{2.43}$$

The computation of the prediction error for (2.39) is complicated.

3. ESTIMATION OF PARAMETERS FOR CURRENT INDIVIDUAL

We consider the same set up of k linear models as given in (2.1)–(2.3) and a $(k + 1)$th linear model of the same type

$$\mathbf{Y}_{k+1} = \mathbf{X}\boldsymbol{\beta}_{k+1} + \boldsymbol{\epsilon}_{k+1}\,. \tag{3.1}$$

Our object is to estimate only the *current parameter* $\mathbf{p}'\boldsymbol{\beta}_{k+1}$ using $\mathbf{Y}_{k+1}$ and the past observations $\mathbf{Y}_1, \cdots, \mathbf{Y}_k$. The estimation of $\mathbf{p}'\boldsymbol{\beta}_i$, $i = 1, \cdots, k$, associated with the past observations $\mathbf{Y}_1, \cdots, \mathbf{Y}_k$ are no longer of interest.

If $\boldsymbol{\beta}_{k+1}$ is considered as a random variable with $E(\boldsymbol{\beta}_{k+1}) = \boldsymbol{\beta}$ and $D(\boldsymbol{\beta}_{k+1}) = \mathbf{F}$, then the Bayes estimator is $\mathbf{p}'\boldsymbol{\beta}_{k+1}^{(b)}$ where, using the formula (2.10),

$$\boldsymbol{\beta}_{k+1}^{(l)} = \boldsymbol{\beta}_{k+1}^{(b)} - \sigma^2\mathbf{U}(\mathbf{F} + \sigma^2\mathbf{U})^{-1}(\boldsymbol{\beta}_{k+1}^{(l)} - \boldsymbol{\beta}). \tag{3.2}$$

When σ^2, $\boldsymbol{\beta}$ and $\mathbf{F}$ are unknown, we substitute their estimates obtained from the past observations $\mathbf{Y}_1, \cdots, \mathbf{Y}_k$ using the formulae (2.16)–(2.18) and compute the empirical Bayes estimator

$$\beta_{k+1}^{(e)} = \beta_{k+1}^{(l)} - c\mathbf{W}\mathbf{B}^{-1}(\beta_{k+1}^{(l)} - \beta_*) \tag{3.3}$$

where $c = (k - m - 2)/(kn - km + 2)$. An alternative is to use the least squares estimator $\beta_{k+1}^{(l)}$ ignoring the previous observations, $\mathbf{Y}_1, \cdots, \mathbf{Y}_k$.

If β_{k+1} is the true value, then the MDE matrix for the empirical Bayes estimator (3.3) is

$$\sigma^2\mathbf{U} - 2c\sigma^2\mathbf{W}\mathbf{U}\mathbf{B}^{-1}\mathbf{U} + c^2\mathbf{W}^2\mathbf{U}\mathbf{B}^{-1}(\delta\delta' + \sigma^2\mathbf{U})\mathbf{B}^{-1}\mathbf{U} \tag{3.4}$$

where $\delta = \beta_{k+1} - \beta_*$, while that for the least square estimator is $\sigma^2\mathbf{U}$. The *expected value* of (3.4) over all possible δ is likely to exceed $\sigma^2\mathbf{U}$ depending on how close the estimates W and $\mathbf{B}^{-1}$ are to their expected values. But for large *individual values* of $\delta\delta'$, the expression (3.4) will exceed $\sigma^2\mathbf{U}$ indicating higher loss for empirical Bayes estimator in individual cases. Such a phenomenon involving higher loss due to underestimation when parameter value is large, and overestimation when parameter value is small, and possible dangers in using Bayes estimators in a routine way were noted in an earlier paper of the author Rao [1974b] while discussing the univariate case.

From a decision maker's viewpoint, the estimator (3.3) is probably the best in the sense of minimising quadratic loss, but for an individual whose 'intrinsic value' is being assessed, the estimator (3.3) may not do full justice, especially when the intrinsic value is high.

For a given estimator $g(\mathbf{Y}_1, \cdots, \mathbf{Y}_{k+1})$ of $\mathbf{p}'\beta_{k+1}$, there are two expressions of loss, one relevant to a decision maker

$$E[(\mathbf{p}'\beta_{k+1} - g)^2 \mid \mathbf{Y}_1, \cdots, \mathbf{Y}_{k+1}] \tag{3.5}$$

and another concerning an individual

$$E[(g - \mathbf{p}'\beta_{k+1})^2 \mid \beta_{k+1}]. \tag{3.6}$$

The decision theory, as developed in statistical literature, ignores the individual aspect and minimises the expression (3.5), thus achieving a minimum over all loss. Perhaps one should see to what extent an individual could be protected by increasing the overall loss. In the present example it is worth examining whether it is possible to choose an alternative estimator g defined by

$$g = \mathbf{p}'\beta_{k+1}^{(e)} \quad \text{if} \quad |\mathbf{p}'\beta_{k+1}^{(l)}| \le d \tag{3.7}$$

$$= \mathbf{p}'\beta_{k+1}^{(l)} \quad \text{if} \quad |\mathbf{p}'\beta_{k+1}^{(l)}| > d \tag{3.8}$$

where d is suitably determined. The consequences of such a decision rule aimed at providing a suitable balance between the expected loss to the decision maker and the loss to an individual can be easily worked out. Some numerical values will be given in a forthcoming paper.

Note 1: In (3.3), we used the estimates of σ^2, β and $\mathbf{F}$ based on past observations $\mathbf{Y}_1, \cdots, \mathbf{Y}_k$. We could, indeed, use the current observation $\mathbf{Y}_{k+1}$ also and obtain updated estimates of σ^2, β, $\mathbf{F}$, using the formulae (2.16)–(2.18) with $k + 1$ instead of k, for substitution of unknown values in (3.2). The empirical Bayes estimator so obtained may be better than (3.3) although the expression for its variance would be complicated.

4. PREDICTION OF A FUTURE OBSERVATION

As in section 3, we have $(k + 1)$ linear models, except that the last component y_n of $\mathbf{Y}_{k+1}$ is not available and has to be estimated (predicted). Let us first suppose that σ^2, β and $\mathbf{F}$ are known and consider the partitioned matrices

$$V = \begin{bmatrix} V_1 & v \\ v & w \end{bmatrix}, \quad Y_{k+1} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \tag{4.1}$$

where V_1 is $(n - 1) \times (n - 1)$ matrix u_1 is the vector of the first $n - 1$ components of Y_{k+1} and X_1 is $(n - 1) \times m$ matrix. Let us define

$$V = (X_1'V_1^{-1}X_1)^{-1}X'_1V_1^{-1}Y_1, \quad U_1 = (X_1'V_1^{-1}X_1)^{-1} \tag{4.2}$$

Theorem 3: *When σ^2, β and F are known, the regression (linear predictor) of y_n on Y_1 is*

$$X_2\beta + (X_1FX_2' + \sigma^2v)'(X_1FX_1' + \sigma^2V_1)^{-1}(Y_1 - X_1\beta) \tag{4.3}$$

and the prediction variance is

$$\sigma^2w + X_2'FX_2 - (X_1FX_2' + \sigma^2v)'(X_1FX_1' + \sigma^2V_1)^{-1}(X_1FX_2' + \sigma^2v). \tag{4.4}$$

The results (4.3) and (4.4) are obtained on standard lines. These expressions do not involve $Y_1, \cdots, Y_k$.

When σ^2, β and F are not known, their estimates can be obtained based on $Y_1, \cdots, Y_k$ as in (2.16)-(2.18) and substituted in (4.3). In such a case the prediction error will be larger than (4.4) depending on how close the estimates of σ^2, β and F are to their true values.

If the problem is simultaneous prediction of the last r components $y_{n-r+1}, \cdots, y_n$ of Y_{k+1}, then we consider partitions as in (4.1) with V_1 as $(n - r) \times (n - r)$ and w as $r \times r$ matrices, u_1 as the vector of the first $(n - r)$ components and u_2 of the last r components of Y_{k+1} and X_1 as $(n - r) \times m$ matrix. With V_1, v, w, X_1 and X_2 so defined, the formula (4.3) provides best predictors of u_2, the vector of last r components of Y_{k+1} and the formula (4.4) represents the mean dispersion error matrix.

ESTIMATION SIMULTANEE DES PARAMETRES DANS DIFFERENTS MODELES LINEAIRES ET APPLICATION A DES PROBLEMES BIOMETRIQUES

RESUME

Une procédure de Bayes empirique est employée dans l'estimation simultanée de vecteurs de paramètres pour un certain nombre de modèles linéaires de Gauss-Markoff. On montre que les estimateurs de Bayes sont meilleurs que ceux des moindres carrés par rapport à la fonction de perte quadratique. En estimant le parametre d'un modèle linéaire particulier, on fait la suggestion de distinguer la perte propre au décideur et celle propre à l'individu. On propose une methode sans l'étudier complétement qui réalise l'équilibre entre ces deux partes. On considere enfin le probleme de la prédiction d'observations futures dans le modèle linéaire.

REFERENCES

Barnard, G. A. [1974]. On ridge regression and general principles of estimation (mimeographed).

Effron, B. and Morris, C. [1972]. Empirical Bayes estimator on vector observations—an extension of Stein's method. *Biometrika, 59*, 335–47.

Effron, B. [1973a]. Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Stat. Assoc. 68*, 117–30.

Effron, B. [1973b]. Combining possibly related estimation problems (with discussion). *J. Roy. Stat. Soc. B, 35*, 379–421.

Effron, B. [1974]. Biased versus unbiased estimation. Tech. Report 1. Stat. Dept. Stanford University.

Fairfield Smith, H. [1936]. A Discriminant function for plant selection. *Ann. of Eugenics.* (London), 7 240–

Geisser, S. [1970]. Bayesian analysis of growth curves. *Sankhya A, 32*, 53–64.

Geisser, S. [1971]. The inferential use of predictive distributions. *Foundation of Statistical Inference*. Holt, Rinehart and Winston, 456–69.

Geisser, S. [1974]. The predictive sample reuse method with applications, (mimeographed).

Hoel, A. E. and Kennard, R. W. [1970a]. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*, 55–68.

Hoel, A. E. [1970b]. Ridge regression: Applications to nonorthogonal problems, *Technometrics, 12*, 69–82.

James, W. and Stein, C. [1961]. Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Stat. and Prob.* University of California press, 361–79.

Lee, J. C. [1972]. On the generalized growth model. *Dissertation* submitted to the university of New York at Buffalo for the Ph.D. degree.

Lee, J. C. and Geisser, S. [1972]. Growth curve prediction. *Sankhya, A34*, 393–412.

Lindley, D. V. and Smith, A. F. M. [1972]. Bayes estimates for the linear model (with discussion). *J. R. Stat. Soc. B, 34*, 1–41.

Panse, V. G. [1946]. An application of the discriminant function for selection in poultry. *J. Genetics* (London) *47*, 242–

Rao, C. Radhakrishna. [1952, 1971]. *Advanced Statistical Methods in Biometric Research*. Hafner, New York.

Rao, C. Radhakrishna. [1953]. Discriminant function for genetic differentiation and selection. *Sankhya, 12*, 229–46.

Rao, C. Radhakrishna. [1965, 1973]. *Linear Statistical Inference and its Applications*. Wiley, New York.

Reo, C. Radhakrishna. [1967]. Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symp. on Math. Stat. and Prob. 1*, 355–72.

Rao, C. Radhakrishna. [1974a]. Some thoughts on regression and prediction—Part I, *Gujarat Statistical Review, 1*, 7–32.

Rao, C. Radhakrishna. [1974b]. Characterization of prior distributions and solution to a compound decision problem. Discussion paper, Indian Statistical Institute, New Delhi.

Smith, A. F. M. [1973]. A general Bayesian model. *J. R. Stat. Soc. B, 35*, 67–75.

Swamy, P. A. V. B. [1970]. Efficient inference in a random coefficient regression model. *Econometrika, 38*, 311–23.

Received February 1975

Key Words: Simultaneous estimation; Compound decision; Ridge regression; Least squares estimation; Empirical Bayes estimation; Prediction of variables.